



UNIVERSIDAD ESAN

FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS

ECONOMÍA Y NEGOCIOS INTERNACIONALES

Los determinantes del rendimiento académico en las regiones del Perú: un enfoque de
Econometría Espacial

Tesis para optar el Título de Licenciada en Economía y Negocios Internacionales,
que presenta:

Yeri Estela Chachico Delgado – DNI: 76924962

Asesor: Eduardo Javier Mantilla Gonzales de la Coterá – DNI: 40381222

LIMA – PERÚ

Noviembre - 2020

Esta tesis

**LOS DETERMINANTES DEL RENDIMIENTO ACADÉMICO EN LAS
REGIONES DEL PERÚ: UN ENFOQUE DE ECONOMETRÍA ESPACIAL**

ha sido aprobada.

Pablo Alfredo Boza Torrejón

Bill William Gee Caballero

María Viviana Prochazka Garavito

Universidad ESAN

Noviembre - 2020

**LOS DETERMINANTES DE RENDIMIENTO ACADÉMICO EN LAS
REGIONES DEL PERÚ: UN ENFOQUE DE ECONOMETRÍA ESPACIAL**

AGRADECIMIENTOS

Dedico esta tesis a mis padres y abuelos por sus consejos,
apoyo incondicional y su paciencia, asimismo
agradezco a todos mis familiares que
son lo mejor y más valioso que
Dios me ha dado.

ÍNDICE

CARÁTULA

PÁGINA DE APROBACIÓN DE TESIS

TÍTULO

DEDICATORIA

ÍNDICE

RESUMEN/ABSTRACT

INTRODUCCIÓN

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	10
1.1 Descripción de la realidad problemática.	10
1.2 Formulación del problema.	12
1.2.1 Problema general.	12
1.2.2 Problema específico.	12
1.3 Objetivos de la investigación.....	13
1.3.1 Objetivo general.	13
1.3.2 Objetivo específico.	13
1.4 Justificación.	13
1.5 Delimitación del Estudio.	13
 CAPITULO II: MARCO TEÓRICO	 14
2.1 Bases Teóricas.	14
2.2 Definición de Variables.....	18
2.3 Antecedentes de la Investigación.....	25
2.4 Contexto de la Investigación.	27
2.5 Hipótesis.....	28
2.5.1 Argumentación.	28
2.5.2 Enunciado de Hipótesis.....	29

CAPITULO III: METODOLOGÍA DE LA INVESTIGACIÓN	30
3.1 Diseño de la Investigación.	30
3.1.1 Econometría Espacial.....	30
3.2 Población y Muestra.	36
3.2.1 Población Objetivo.	36
3.2.2 Método de muestreo.....	36
3.2.3 Tamaño de la muestra.	37
3.3 Método de Recolección de Datos.	37
3.4 Método de Análisis de Datos.....	38
3.5 Técnicas para el Procedimiento y Análisis de la Información.	39
 CAPITULO IV: ANÁLISIS DE LA INTERPRETACIÓN DE LAS PRUEBAS ESTADÍSTICAS, FÓRMULAS, RESULTADOS	 40
4.1 Modelo de regresión general.	40
4.1.1 Análisis Exploratorio de Datos Espaciales.	40
4.1.2 Modelación Espacial del modelo general.	46
4.2 Modelo de regresión según ámbito geográfico.	52
4.2.1 Análisis Exploratorio de Datos Espaciales.	52
4.2.2 Modelación Espacial según ámbito geográfico.	54
4.2.3 Análisis de los Resultados según ámbito geográfico.....	67
 CAPITULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES	 68
BIBLIOGRAFÍA	70
ANEXOS.....	78

RESUMEN

La literatura nacional manifiesta poca evidencia sobre el uso de la econometría espacial como herramienta para estudiar el rendimiento académico y sus determinantes; por ello el objetivo de la investigación es identificar los determinantes del rendimiento académico en Comprensión Lectora en las regiones geográficas del Perú e investigar la existencia de autocorrelación espacial. La metodología usada en la presente investigación para el modelo general es el Modelo Espacial de Error (SEM) de efectos fijos en panel data, cuyos resultados indican que las variables de servicios básicos, educación inicial y tamaño promedio de clase tienen estadísticos significativos. Por otro lado; para la Costa las variables que resultaron estadísticamente significativos fueron el acceso a Internet y la educación inicial, para la Sierra fueron el acceso a Internet, la educación inicial y el tamaño promedio de clase; mientras que para la Selva fueron las variables de educación inicial y las Necesidades Básicas Insatisfechas. Con estos resultados se llega a la conclusión de la existencia de autocorrelación espacial del rendimiento académico, agrupándose por ubicación geográfica. Asimismo, se resalta la importancia que tiene la educación inicial sobre el rendimiento académico, por ello el Gobierno debería implementar políticas públicas que permita aumentar la cobertura y calidad de la educación inicial para todas las regiones.

Palabras Clave: Dependencia espacial, Rendimiento Académico, Acceso a Internet, Educación Inicial, Tamaño de Clase, Necesidades Básicas Insatisfechas.

ABSTRACT

The national literature shows little evidence on the use of spatial econometrics as a tool to study academic performance and its determinants; for this reason, the objective of the research is to identify the determinants of the academic performance in Reading Comprehension in the geographical regions of Peru and to investigate the existence of spatial autocorrelation. The methodology used in the present investigation for the general model is the Spatial Error Model (SEM) of fixed effects in panel data, whose results indicate that the variables of basic services, early childhood education and average class size have significant statistics. On the other hand; for the Coast, the variables that were statistically significant were Internet access and Early childhood education; for the Highland, they were Internet access, Early childhood education, and average class size; while for the Jungle they were the variables of Early childhood education and Unsatisfied Basic Needs. With these results, it is concluded that there is a spatial autocorrelation of academic performance, grouped by geographic location. Likewise, the importance of Early childhood education on academic performance is highlighted; therefore, the Government should implement public policies to increase coverage and quality of Early childhood education for all regions.

Key Words: Spatial dependency, Academic Performance, Internet access, Early Childhood Education, Class Size, Unsatisfied Basic Needs.

INTRODUCCIÓN

La educación es un factor que influye notablemente al crecimiento económico de las regiones debido a los efectos positivos que genera el conocimiento para la sociedad. Los resultados de las pruebas PISA, evidencian que el Perú enfrenta un problema de baja calidad educativa, pues los rendimientos académicos de los alumnos son bajos a nivel mundial.

Estudiar el rendimiento académico de los alumnos es complejo debido a que esta es una variable que tiene muchos posibles factores que la expliquen. La mayoría de estudios que se han enfocado en analizar los determinantes del rendimiento en el Perú, han estimado modelos explicativos generales sin considerar modelos diferenciados para distintas subpoblaciones. Asimismo, son pocos los estudios que existen sobre el rendimiento académico, utilizando un enfoque espacial. Por ello, el objetivo de la presente investigación es identificar los determinantes del rendimiento académico en Comprensión Lectora según el ámbito geográfico y determinar la existencia de autocorrelación espacial del rendimiento académico en las regiones del Perú.

El documento se estructura de la siguiente manera. En el Capítulo I se formula el planteamiento del problema. En el Capítulo II se revisa el marco teórico de la investigación. En el Capítulo III se explica la metodología a seguir. En el Capítulo IV se explica el tratamiento de la información desde un enfoque espacial y los resultados preliminares de los determinantes del rendimiento académico en Comprensión Lectora, a nivel general y según ámbito geográfico. En el Capítulo V se muestran las conclusiones. Finalmente, se muestran las referencias y los anexos del trabajo.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática.

“La educación es un pilar fundamental para el desarrollo de los países; que tiene implicancias a nivel individual, por el lado de los ingresos y a nivel global contribuye con un mayor progreso para los países en términos de crecimiento económico”, pues las personas con mayor educación tienen niveles de productividad más altos y por ende contribuyen al crecimiento económico (Zuniga, Ordoñez, Ordoñez, & Martínez, 2018, p. 3).

El capital humano es un factor explicativo de la teoría del crecimiento económico, dónde se ve cómo la inversión en educación genera beneficios a futuro. Por eso, durante varios años “los gobiernos, tanto de países desarrollados como de países en vía de desarrollo, han centrado su atención en aportar una mayor proporción del gasto hacia la educación” (Palacios, 2009, p. 4).

El Perú es uno de los países que tiende a destinar un bajo porcentaje del PBI al sector educativo. Como se puede observar en el *Anexo N.º 7.1*, aunque el Perú haya incrementado su asignación del gasto público en educación al 4% del PBI en el año 2015, aún está por debajo comparado con los demás países como lo señalan los indicadores de desarrollo del Banco Mundial (BANCO MUNDIAL, 2018).

Sin embargo, cuando se habla de gasto público en educación, el gastar una mayor cantidad de dinero no necesariamente se traduce en mayor calidad educativa y por lo tanto en un mejor rendimiento académico. En la Tabla N.º 1 se puede observar el caso de Perú, que aun incrementando recursos hacia la educación pública, los resultados de las evaluaciones internacionales evidencian que el Perú aún tiene un bajo rendimiento académico en Comprensión Lectora comparado con los países de Latinoamérica (MINEDU, 2018).

**Tabla N° 1. Variación de los resultados en Comprensión Lectora -
Latinoamérica (2009 – 2018)**

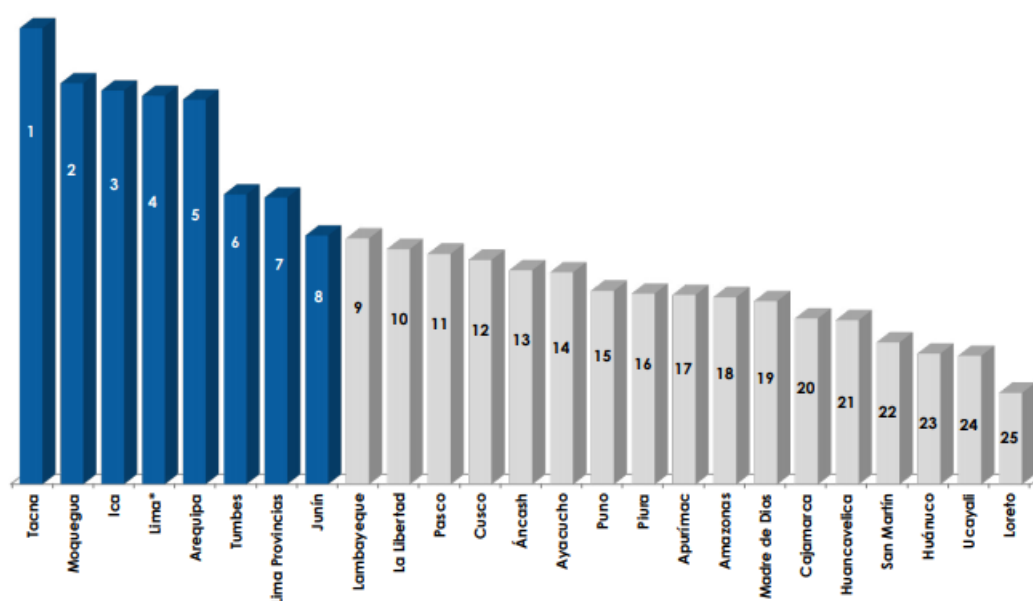
<i>Países</i>	<i>2009</i>	<i>2012</i>	<i>2015</i>	<i>2018</i>	<i>Tendencia promedio</i>
<i>Panamá</i>	-	-	-	377	-
<i>República Dominicana</i>	-	-	358	342	-16,1
<i>Costa Rica</i>	443	441	427	426	-5,3
<i>México</i>	425	424	423	420	-1,7
<i>Colombia</i>	413	403	425	412	-0,3
<i>Uruguay</i>	426	411	437	427	+0,3
<i>Brasil</i>	412	407	407	413	+0,3
<i>Chile</i>	449	441	459	452	+1
<i>Argentina</i>	398	396	-	402	+2
<i>Perú</i>	370	384	398	401	+10,3

Fuente: Ministerio de Educación - Elaboración: Propia

Debido a la naturaleza multifactorial del rendimiento académico, la literatura propone un marco conceptual a partir de la función de producción que incorpora una serie de variables que explican los resultados de los rendimientos académicos como variables asociados al colegio y variables que están relacionados con las características de los estudiantes y sus hogares (Glewwe & Kremer, 2006; Beltrán & Seinfeld, 2012).

A nivel nacional, el Índice de Competitividad Regional (INCORE) mide el desempeño de las regiones del país a través de seis pilares de competitividad, siendo uno de ellos la educación. De esta manera en la Figura N.º 1, se puede observar que las regiones de la Costa como Lima, Moquegua, Arequipa y Tacna muestran mayor competitividad regional en educación que las regiones de la Sierra y Selva (Cueto, 2007).

Figura N.º 1: Índice de Competitividad Regional en Educación



Fuente: IPE (2018)

Dado que el país muestra desigualdades regionales del rendimiento escolar es de importancia estudiar el porqué de estas diferencias. En el Perú existen pocos trabajos de investigación que analizan el rendimiento académico desde un enfoque espacial. Por ello, en el presente trabajo se plantea identificar los determinantes del rendimiento académico en las regiones del Perú tomando como base la función de producción de educación y determinar la existencia de dependencia espacial en el rendimiento académico, pues este permite analizar cómo afecta el rendimiento académico de una región sobre las regiones que tienen una proximidad geográfica.

1.2 Formulación del problema.

1.2.1 Problema general.

¿Existe autocorrelación espacial del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 en las regiones del Perú?

1.2.2 Problema específico.

¿Cuáles son los determinantes del rendimiento académico en Comprensión Lectora según el ámbito geográfico, durante los años 2012 a 2016?

1.3 Objetivos de la investigación.

1.3.1 Objetivo general.

Determinar la existencia de autocorrelación espacial del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 en las regiones del Perú.

1.3.2 Objetivo específico.

Identificar los determinantes del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016, según el ámbito geográfico.

1.4 Justificación.

La presente investigación propone estudiar el rendimiento académico desde un enfoque espacial, con el objetivo de determinar la existencia de dependencia espacial e identificar los determinantes del rendimiento académico en las regiones geográficas del Perú. De esta manera, el aporte principal de los resultados de la investigación es difundir la importancia que tiene el saber en qué determinantes del rendimiento académico invertir en cada una de las regiones geográficas. Así, el Estado puede ajustar las políticas públicas en educación y hacerlas más efectivas en aquellas regiones que más lo necesiten.

1.5 Delimitación del Estudio.

El presente tema de investigación delimita el estudio a identificar los determinantes del rendimiento académico en las regiones geográficas y determinar la existencia de dependencia espacial en esta variable, a través de la econometría espacial durante los años 2012 a 2016, para una muestra de estudiantes exclusivamente de segundo grado de primaria.

CAPITULO II: MARCO TEÓRICO

Los modelos de crecimiento económico consideran al capital humano como un factor de producción importante, pues los individuos que poseen mayor educación son más productivos y tienen mayor capacidad para adaptarse a las nuevas tecnologías e innovar procesos. Desde este enfoque, la educación en las escuelas no sólo es importante para la generación de conocimientos sino también para el desarrollo de habilidades de los estudiantes, que finalmente en la función de producción de la educación son medidos en términos de rendimiento académico.

En esta sección se presentarán los modelos de crecimiento económico desarrollados por Denison (1962), Becker (1964), Uzawa (1965) & Lucas (1988) y Mankiw, Romer & Weil (1992) con el objetivo de entender el concepto de capital humano y su importancia en la función de producción de la educación.

2.1 Bases Teóricas.

2.1.1 Modelos de crecimiento económico.

Los estudios realizados por Denison (1962) Becker (1964), Uzawa (1965) & Lucas (1988) y Mankiw, Romer & Weil (1992) han promovido el desarrollo del marco teórico y análisis empírico de la educación desde una perspectiva económica. Gracias a ellos surge la Economía de la Educación, que tiene una doble finalidad: por un lado, estudia los efectos de la educación como factor de crecimiento económico y, por otro, analiza los aspectos económicos de los procesos educativos.

En 1950 se empieza a notar la dificultad de explicar el crecimiento económico, por lo cual se comienza a estudiar el papel que desempeñan los trabajadores en el medio económico y la trascendencia de su conocimiento y el *know - how* que tienen estos para la producción. Por ello, se da comienzo a estudiar al capital humano como factor determinante para el crecimiento económico (Cardona, Montes, Vásques, Villegas, & Brito, 2007).

Para Schultz (1961), la teoría del crecimiento económico sustenta al capital humano como formas de inversión que producen beneficios a futuro con mayores ingresos para las personas con educación, y por ende a la sociedad y al crecimiento económico de una región. De esta manera, los seres humanos aumentan sus posibilidades de trabajar en mejores puestos e incrementar su calidad de vida.

Por otro lado, Deninson (1962) señala que cualquiera que sea el periodo que se analice, el crecimiento económico resulta de un incremento de la fuerza de trabajo, educación, capital y avance de los conocimientos. De tal forma se llega a la conclusión que la relación educación-crecimiento económico es directa, es decir, que a mayor inversión en la educación habrá un mayor crecimiento económico.

Asimismo, Becker (1964) considera al sector educación como principal productor del capital humano, ya que genera conocimientos y desarrolla habilidades del estudiante mediante el cual tendrá la posibilidad de obtener salarios más elevados e incrementar el ingreso per cápita de la sociedad. Sin embargo, es importante mencionar también, que la productividad de los trabajadores no solo depende del capital humano sino también de otros factores como el esfuerzo y la motivación de estos.

En los años ochenta, Lucas (1988) retoma el trabajo de Uzawa (1965) con el objetivo de determinar los componentes que explican la creación del capital humano, llegando a la conclusión que el crecimiento del capital humano está relacionado básicamente con dos factores: la calidad de la educación y el porcentaje de tiempo que las personas dedican al estudio. Asimismo; el modelo desarrollado por Mankiw, Romer y Weil (1992) constituye uno de los modelos de crecimiento empíricos más notables para explicar el crecimiento de una economía, pues se afirma que no solo el capital físico incide sobre el crecimiento de una región sino también el capital humano.

En resumen; los modelos de crecimiento económico demuestran que la inversión en capital humano genera una mayor productividad de los trabajadores, ya que una mano de obra más capacitada utiliza el capital de manera más eficiente. De este modo, el retorno de inversión del capital humano beneficia a la sociedad incrementando el crecimiento de la región (Elías, 2004).

2.1.1.1 Capital Humano.

Schultz (1961) y Becker (1964) definen el capital humano como el nivel de habilidades y conocimientos incorporados en el individuo gracias a la educación. Estos autores sostienen que la acumulación de capital humano puede ser vista como inversión, ya que al aumentar su productividad, aumenta también sus ingresos futuros tanto del individuo como de la nación en conjunto.

En el caso de los niños, el desarrollo de su capital humano, si bien depende de la educación que se le da en el colegio también depende en una medida importante de sus hogares. Por ello, se hace uso de la función de producción para conocer los determinantes del rendimiento académico.

2.1.1.2 Función de producción de la educación.

La función de producción general muestra la relación entre la cantidad de insumos y factores intervinientes para producir un determinado bien, tomando en consideración la calidad del mismo. Para el sector educación, la función de producción se mide a través de cómo la escuela utiliza los recursos que tiene /input/ para producir un bien educativo como salida /output/, que puede ser expresado en términos de rendimiento académico (Hanushek, 1989).

Si la función de producción fuera conocida, sería posible predecir qué pasaría si la cantidad de recursos es cambiada, sin embargo esta no es conocida, por ello debe ser inferida empleando información de los estudiantes y sus escuelas por medio de estimaciones. Por eso, los economistas sugieren una función de producción en educación (Rajimon, 2010).

Si bien el contenido de la presente investigación es determinar la existencia de dependencia espacial e identificar los determinantes del rendimiento académico, es de importancia presentar diferentes modelos de función de producción del rendimiento empleados por diversos autores para conocer cómo se ha venido relacionando las entradas y salidas en el proceso del logro académico.

La función de producción en educación tuvo origen en un estudio realizado por Coleman (1966) en el cual llega a la conclusión de que la diferencia de la cantidad de recursos asignados en las escuelas tiene pequeños efectos en el rendimiento y confirma que los antecedentes familiares y las características de los estudiantes tienen una relación más significativa con el rendimiento académico.

Por otro lado, Hanushek (1989), afirma que las escuelas sí hacen la diferencia en el desempeño educativo, a su vez argumenta que las variables explicativas más utilizadas en la función de producción de la educación son los factores familiares, los factores que miden el desarrollo de una región y los factores escolares como las características de los docentes, el tamaño de clase e instalaciones.

Asimismo, Beltrán y Seinfeld (2012) categoriza los determinantes del rendimiento académico por el lado de la demanda, que están relacionados con las características de los estudiantes y sus hogares, y por el lado de la oferta, aquellos asociados al colegio, en el cual se incluye variables relacionadas con la gestión de la educación, variables que reflejan la calidad de la infraestructura escolar y variables que miden la calidad de los docentes.

Las evidencias sugieren que el rendimiento escolar depende de factores socioeconómicos, de la calidad del docente, de las condiciones de la escuela y de las características del alumno. Basado en el modelo de la función de producción de la educación utilizado por Glewwe & Kremer (2006), utilizaremos la siguiente expresión:

$$A_{it} = f(Q_{it}, C_{it}, H_{it}, I_{it}),$$

donde A_{it} es el puntaje promedio del rendimiento académico en la región i durante el año t , Q_{it} es el vector con las variables asociadas a la escuela y características de los docentes en la región i durante el año t , C_{it} es el vector de características del alumno en la región i durante el año t , H_{it} es el vector de características del hogar en la región i durante el año t e I_{it} es el vector de insumos educativos que los padres pueden controlar, como la compra de útiles escolares en la región i durante el año t .

2.2 Definición de Variables.

2.2.1 Rendimiento Académico.

El rendimiento académico es el resultado de la evaluación del conocimiento de lo que una persona ha aprendido después de un proceso de formación. Asimismo, es definido también como un valor atribuido al logro del estudiante, el cual habitualmente es expresado por medio de calificaciones escolares (Martínez-Otero, 2007).

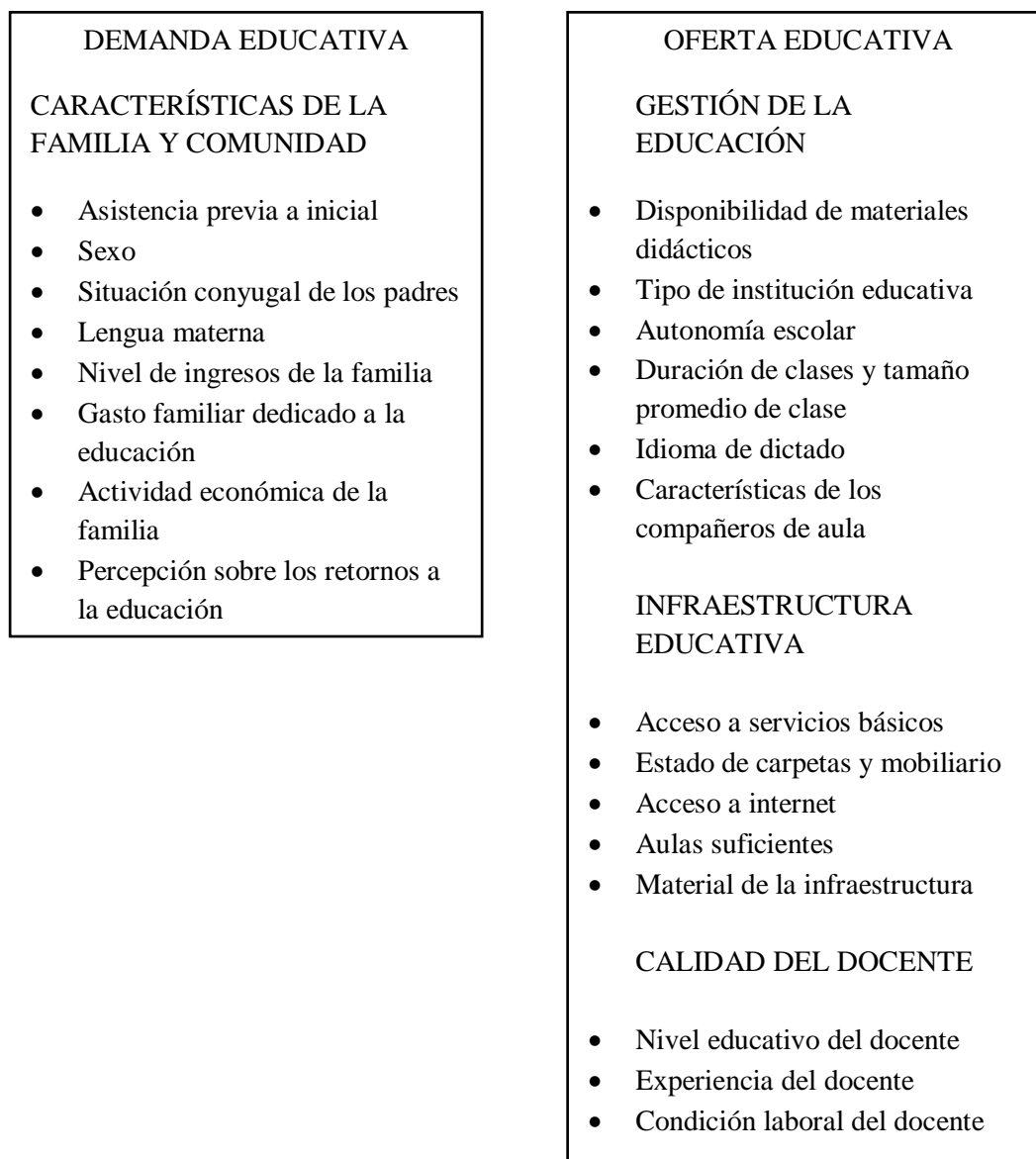
El objetivo del rendimiento académico es medir el nivel de aprendizaje alcanzado por el alumno en diversas materias por medio de puntuaciones, que tienen una relación con la superación o no de determinadas pruebas. (Caballero, Abello, & Palacio, 2007). De este modo, es importante precisar que en el trabajo de investigación se utilizará el rendimiento académico en Comprensión Lectora como variable proxy del rendimiento académico.

En el Perú, la principal evaluación del Ministerio de Educación es la ECE que se aplica desde el año 2007 a los estudiantes de segundo grado de primaria en las asignaturas de Matemática y Comprensión Lectora, con el fin de realizar un seguimiento continuo del rendimiento educativo de los estudiantes permitiendo evaluar así su evolución.

2.2.1.1 *Determinantes del rendimiento académico.*

Las variables determinantes del rendimiento académico se dividen en dos categorías, uno por el lado de la oferta y el otro por lado de la demanda educativa. Ambas perspectivas son relevantes para el desarrollo de los estudiantes por lo cual se complementan para determinar los resultados educativos (Filmer, 2003)

Figura N° 2: Principales determinantes del rendimiento educativo en el Perú



Fuente: Beltrán y Seinfeld (2012)

2.2.1.1.1 Demanda Educativa.

Las variables determinantes de la demanda educativa están vinculadas con las características de los grupos sociales que condicionan la efectividad de la educación. En este apartado se muestran las características de los individuos y de su entorno familiar y social que afectan en cierto modo a los resultados educativos de los estudiantes (Beltrán & Seinfeld, 2012).

a. Características de la familia y comunidad

La educación de todo ser humano comienza en la familia y se extiende al centro educativo, por ello es necesario el conocimiento de las características de la familia y la comunidad en la que viven los escolares (Sánchez, 2012).

- ***Asistencia a educación inicial.***

La educación en los primeros años de vida del individuo es muy esencial para el desarrollo del mismo, pues no solo tiene efectos positivos en el corto plazo, sino que también tiene efectos sociales y económicos a largo plazo. De este modo, el mejor rendimiento que se alcanza origina un círculo virtuoso de desarrollo educativo favoreciendo la igualdad de oportunidades (Escobar, 2006).

Las investigaciones de Currie & Thomas (2000), Myers (1992) y Berlinski, Galiani & Gertler (2006) concluyen que la asistencia a la educación inicial fomenta un mejor rendimiento a futuro. Por su parte Barnett & Esposito (2006) también encuentran en su estudio que los niños que han asistido por dos o más años a la educación inicial tiene un efecto positivo sobre las habilidades para el uso del lenguaje.

Como se puede observar, diversos autores coinciden en afirmar que la base educativa que posee un estudiante es uno de los principales factores del rendimiento académico. Debido a la importancia que tiene esta variable, se busca determinar la relación que guarda esta variable con el rendimiento académico.

- ***Necesidades Básicas Insatisfechas.***

Las Necesidades Básicas Insatisfechas es un método utilizado para medir la pobreza en América Latina, ya que esta influye en todos los contextos socioculturales en los que se desenvuelven los individuos. Por lo mismo, se da principal atención a la educación, pues una buena inversión en educación tiene efectos positivos en la productividad laboral y por ende en los ingresos de las personas, que se traduce en la reducción de pobreza de la región o país. (Rodríguez, 2015).

Bajo el método de las Necesidades Básicas Insatisfechas, se elige una serie de indicadores censales que permiten constatar si los hogares satisfacen o no algunas de sus necesidades principales como viviendas con características físicas inadecuadas, con hacinamiento y sin servicios higiénicos, con niños que no asisten a la escuela y con alta dependencia económica (MEF, s.f.).

Los estudios de Jadue (1997) y (Otero, 2006) concuerdan que el rendimiento académico de un alumno depende de un trabajo conjunto de escuela y familia. Llegando a la conclusión que los alumnos de estratos sociales más bajos muestran menores rendimientos en comparación con estudiantes de estratos sociales más altos, sin embargo existe excepciones dónde el factor motivación juega un rol importante en el rendimiento académico del alumno. (Ramírez, Quiñones, Erasmo, Salazar, & Aníbal, 2011).

2.2.1.1.2 Oferta Educativa.

Las variables determinantes por el lado de la oferta son aquellos que se refieren a la calidad y cobertura de los servicios donde se produce el proceso educativo. Esta categoría se divide en tres grupos: primero, las variables relacionadas con la gestión de la educación; segundo, las variables que muestran la calidad de la infraestructura escolar y por último, las variables que miden la calidad de los docentes (Beltrán & Seinfeld, 2012).

a. Gestión de la Educación.

La gestión educativa está formada por un grupo de procesos estructurados que permiten que una institución educativa logre sus metas y objetivos. La función esencial de la gestión es desarrollar una educación de calidad lo que se entiende como la conducción del sistema educativo hacia el logro de procedimientos educativos de calidad (Correa, Correa, & Alvarez, s.f.).

La calidad de educación está medida por diferentes variables; sin embargo, para el análisis de esta investigación se tomará el tamaño de clase como una medida de calidad educativa pues se tiene información de este indicador.

- ***Tamaño promedio de clase.***

El tamaño de clase es uno de los aspectos más debatidos en educación, ya que existe una percepción generalizada de que las clases más pequeñas permiten al profesor dedicar más tiempo a cada alumno y menos tiempo a la gestión general del aula, dando una mejor enseñanza adaptada a las necesidades individuales de cada alumno, lo que garantiza un mejor rendimiento (Ministerio de Educación, Cultura y Deporte, 2016).

Uno de los experimentos más famosos en este campo es el experimento STAR (Student Teacher Achievement Ratio), realizado en los años 80 en Tennessee para investigar si la reducción en el tamaño de las clases tiene una relación inversa con el rendimiento de los estudiantes. Para poder llevar a cabo el estudio, se dividió a los estudiantes en clases de 15 alumnos y de 23 alumnos, dando como resultado que el rendimiento de los estudiantes mejoro en los estudiantes que estaban en clases pequeñas. Con lo cual, llegan a la conclusión de que el tamaño de clases sí influye en el rendimiento académico de los alumnos (Mosteller, 1995).

b. Infraestructura Educativa.

La infraestructura educativa es el conjunto de todos los elementos que conforman el espacio físico dónde se realiza los procesos de enseñanza y aprendizaje. La infraestructura del colegio es un factor determinante para el rendimiento escolar pues cumple un rol motivacional y funcional; es decir, produce una mejor disposición de los estudiantes hacia el aprendizaje y favorece el proceso de enseñanza (Campana, Velasco, Aguirre, & Guerrero, 2014).

Beltrán & Seinfeld (2012) dividen la infraestructura educativa en distintos factores como el acceso a servicios básicos, el mobiliario, el acceso a Internet, entre otros. A continuación, se detallan cada uno de ellos y se verá la relación que tienen estos con el rendimiento académico.

- ***Acceso a Servicios Básicos.***

Los servicios básicos son aquellos servicios como el agua, el desagüe y la energía eléctrica que poseen los planteles escolares; los cuales son importantes para brindar una educación de calidad a los escolares. Asimismo, el Ministerio de educación afirma que el contar con acceso a los tres servicios básicos contribuye positivamente al bienestar de la comunidad educativa (MINEDU, 2017a).

Con respecto al acceso a la energía eléctrica, una mayor iluminación hace que los estudiantes aprovechen más el tiempo dando como resultado que la tasa de alfabetización aumente en el país. Además, el contar con energía eléctrica permite que los estudiantes tengan acceso a Internet, con lo cual obtienen información y conocimiento (Niu, Jia, Wang, He, Hu, & Liu, 2013).

Con respecto al acceso de agua potable y de saneamiento la escasez del agua y el saneamiento deficiente son perjudiciales para la salud de los niños en edad escolar. Las enfermedades relacionadas con el agua y el saneamiento tienen efecto en la asistencia de la escuela, ya que los alumnos no pueden asistir a la escuela cuando están delicados de salud, lo que a su vez tiene un efecto sobre el rendimiento académico (UNICEF, 2017).

- ***Mobiliario escolar.***

Uno de los indicadores para medir el cumplimiento de las obligaciones del Estado en materia de educación, de acuerdo con el modelo de Tomasevski (2004), es la asequibilidad de la educación. Este indicador supone que las escuelas han de contar con el mobiliario y el equipo indispensables (mesas, sillas, pizarrones) para el logro de los propósitos educativos. Sin embargo, no basta con conocer si estos recursos escolares existen o no, sino que además importa si estos son suficientes y se encuentran en las condiciones adecuadas (INEE, 2014).

El disponer de mobiliario escolar adecuado y suficiente apoya a los estudiantes a recibir los nuevos conocimientos con mayor facilidad. De igual modo, permite desarrollar las actividades de aprendizaje en ambientes cómodos que facilitan la participación de los estudiantes durante las clases, lo cual tiene efecto positivo en los rendimientos académicos (Hernández, 2011).

- ***Acceso a Internet.***

El acceso a Internet puede mejorar la calidad de la educación, ya que brinda acceso hacia una gran cantidad de información, conocimiento y recursos educativos, aumentando las oportunidades de aprendizaje. Sin embargo, se debe tener en cuenta que contar con acceso a Internet y encontrar información es una cosa, pero comprenderla y darle buen uso a la información es otra cuestión, que requiere de la intervención de instituciones educativas si se quiere estar en condiciones de explotar las capacidades potenciales del Internet al máximo (Internet Society, 2017).

c. Calidad del Docente.

La eficacia de los docentes es primordial para alcanzar una educación de calidad. Desde ese punto de vista, una de las prioridades de los sistemas educativos debería ser el mantener e incrementar la calidad de los profesores (UNESCO, 2007).

- ***Nivel Educativo del Docente.***

Uno de los factores que aproxima la calidad de los maestros es si estos poseen una certificación o no.

En el informe de McKinsey, se demuestra que los niños bajo la enseñanza de docentes sin calificación adecuada cuentan con escasas probabilidades de recuperarse en la escuela en los años siguientes, por lo cual el docente es un factor decisivo para la calidad educativa. Asimismo, en este informe se encuentra que la variable de porcentaje de profesores titulados tiene efectos positivos sobre el rendimiento de los estudiantes (Fullan, Schleicher, Lee, Gopinathan, & Hill, 2007).

2.3 Antecedentes de la Investigación.

Existen diversos trabajos en el Perú que buscan identificar los determinantes del rendimiento académico, sin embargo son pocas las investigaciones que existen sobre el rendimiento académico utilizando un enfoque espacial.

El trabajo realizado por Fuentes, Tarazona, Tarquino & Villanueva (2016) analiza, evalúa y propone un modelo de indicadores de eficiencia para la educación primaria, con el objetivo de determinar las variables que tienen mayor impacto sobre los logros de aprendizaje de los estudiantes en Comprensión Lectora. Los resultados del análisis indican que la educación inicial, los alumnos por docente y los profesores titulados en primaria son variables significativas en los resultados académicos, por ello se concluye que los determinantes que más afectan al rendimiento son estas variables (Fuentes, Tarazona, Tarquino, & Villanueva, 2016).

Asimismo, Beltrán & Seinfeld (2012) analizan los principales factores que influyen sobre el rendimiento académico de los alumnos; con el objetivo de identificar los determinantes del rendimiento diferenciando aquellos que provienen del lado de la oferta educativa y los que se derivan de la demanda del servicio educativo. Los resultados del análisis indican que los indicadores vinculados con la infraestructura y la educación inicial son variables significativas en los resultados académicos. Además la heterogeneidad del impacto sobre el desempeño se explica, principalmente, por situaciones de desigualdad entre grupos para acceder a una oferta educativa adecuada. En conclusión se destaca la importancia que tiene la educación inicial y la infraestructura educativa en el futuro desempeño escolar de los alumnos (Beltrán & Seinfeld, 2012).

Introduciendo la metodología espacial el trabajo realizado por Naidoo, Eeden & Munch (2014) analiza la relación entre las variables socioeconómicas de la región y el rendimiento escolar con el objetivo de identificar las variables explicativas más relevantes para el análisis espacial del rendimiento académico. Los hallazgos clave de esta investigación sugieren que las escuelas de Ciudad del Cabo están agrupadas en términos de rendimiento escolar y las variables que influyen en esta son las características del hogar. En conclusión, las ubicaciones físicas de las escuelas secundarias en el municipio de Ciudad del Cabo están agrupadas y no distribuidas al azar dentro del contexto sudafricano (Naidoo, Eeden, & Munch, 2014).

Asimismo, se destaca el trabajo realizado por Hincapié (2017), que analiza los efectos contagio del gasto público sobre el rendimiento educativo para el departamento de Antioquia con el propósito de estimar la relación entre estas dos variables y determinar si el rendimiento educativo presenta dependencia espacial utilizando herramientas de la econometría espacial. Los resultados del trabajo revelan que el rendimiento académico en educación básica de los municipios de Antioquia presenta una dependencia espacial positiva. En conclusión, se demuestra que el rendimiento académico de un municipio si depende del rendimiento educativo de sus vecinos tanto en Comprensión Lectora como en Matemática (Hincapié, 2017).

Por otro lado, a nivel nacional se resalta el trabajo de Elías & Rey (2011), que analizan la convergencia espacial del rendimiento educativo en las provincias del Perú, con el objetivo de determinar si existe dependencia espacial en esta variable. Los resultados de dicho análisis muestran las diferencias del rendimiento académico en las regiones Costa, Sierra y Selva, pues las regiones costeras tienden a tener un mayor rendimiento académico que los estudiantes de la Sierra y Selva. En conclusión, se demuestra que existe dependencia espacial del rendimiento académico, es decir que una región depende del rendimiento educativo de sus regiones colindantes.

2.4 Contexto de la Investigación.

La educación es fundamental para impulsar el crecimiento de una región, pues influye positivamente en la calidad de vida de las personas y en el crecimiento económico de las regiones, debido al aumento en la productividad de los individuos. Por ello, Hanushek & Wößmann(2007) consideran que el aprendizaje es un determinante del crecimiento económico, a través de su impacto en el capital humano.

Por otro lado, en el Perú la educación es considerada un derecho fundamental para todos los individuos; por ello, el Gobierno asume la responsabilidad de proveer gratuitamente este servicio a toda la ciudadanía en todos sus niveles y modalidades, favoreciendo así la ampliación de la cobertura de esta. De este modo, el Estado se encarga de supervisar la calidad de la provisión de servicios educativos en las Instituciones de gestión pública y, en el caso de las Instituciones de gestión privada, las regula y supervisa (MINEDU, 2016a).

A nivel nacional, según el Índice de Competitividad Regional (INCORE) las regiones con mayor crecimiento económico y social se encuentran en la Costa, mientras las menos competitivas se encuentran en la Sierra y Selva. De igual forma, el Instituto Nacional de Estadística e Informática (2018), sostiene que existe un comportamiento diferenciado según el ámbito geográfico en el que se encuentran los estudiantes. En las regiones de la Costa, se concentran en mayor medida las familias que tienen más acceso a la infraestructura escolar mientras que en la Sierra y Selva se reduce considerablemente. Asimismo, la región Costa tiene mayor tasa de asistencia a educación inicial, como mayor porcentaje de profesores titulados que las regiones de la Sierra y Selva.

En el marco de estas diferencias regionales, es necesario evaluar el rendimiento académico durante los años del 2012 a 2016 a través de la Evaluación Censal de Estudiantes (ECE) con el objetivo de identificar los determinantes del rendimiento académico de los estudiantes según su ámbito geográfico y determinar la existencia de dependencia espacial en esta variable.

2.5 Hipótesis.

2.5.1 Argumentación.

De acuerdo a la primera ley de la geografía de Tobler (1970), las regiones cercanas están más relacionadas que las lejanas; por ello los resultados del rendimiento académico deberían de ser similares en las regiones cercanas. Esta relación ha sido contrastada empíricamente por diversos autores - como Dall'erba & Le Gallo (2004); Elías & Rey (2011) e Hincapié (2017) - que analizan el rendimiento educativo de países y provincias manteniendo una perspectiva espacial encontrando evidencia empírica que demuestra que el rendimiento académico de una unidad de análisis si depende del rendimiento educativo de sus vecinos.

H1: De igual forma, se espera que exista autocorrelación espacial positiva del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016.

Por otro lado, basándose en el modelo de la función de producción de la educación, el rendimiento escolar depende de factores socioeconómicos, de la calidad del docente, de las condiciones de la escuela y de las características del alumno (Glewwe & Kremer 2006). Así, estos factores se convierten en elementos determinantes del rendimiento académico, los cuales han sido contrastados empíricamente por diversos autores en diferentes contextos (Coleman, 1966; Hanushek 1989; Naidoo, Eeden, & Munch, 2014; Beltrán y Seinfeld, 2012).

La mayoría de estudios se han enfocado en analizar el rendimiento académico sobre la base de modelos explicativos generales sin considerar modelos diferenciados para distintas subpoblaciones (Fuentes, Tarazona, Tarquino & Villanueva, 2016; Beltrán & Seinfeld 2012). Si bien los modelos a nivel nacional brindan una tendencia de los rendimientos académicos, ello no permite tener herramientas para tomar decisiones focalizadas. A partir de las evaluaciones ECE, se encuentra diferencias en los resultados del rendimiento académico en las regiones Costa, Sierra y Selva, siendo contrastado empíricamente por Elías & Rey (2011).

H2: Por esta razón, se espera que los determinantes del rendimiento académico en Comprensión Lectora sean diferentes según el ámbito geográfico, durante los años 2012 a 2016.

2.5.2 Enunciado de Hipótesis.

2.5.2.1 *Hipótesis General.*

Existe autocorrelación espacial positiva del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 en las regiones del Perú.

2.5.2.2 *Hipótesis Específica.*

Los determinantes del rendimiento académico en Comprensión Lectora durante los años del 2012 a 2016 son diferentes, según el ámbito geográfico.

CAPITULO III: METODOLOGÍA DE LA INVESTIGACIÓN

3.1 Diseño de la Investigación.

La investigación desarrolla un diseño no experimental, ya que las variables no son manipuladas ni controladas, asimismo se observa los fenómenos tal y como se dan en su contexto natural para después ser analizados. Es cuantitativo, pues se analizan las variables con los datos disponibles de ESCALE y de los Resultados de la Evaluación Censal de los Estudiantes. Con estos datos se construye una base de datos de tipo panel data, que permite trabajar con un mayor número de observaciones y captura la heterogeneidad entre las regiones. Por último, es de tipo descriptivo pues se busca identificar los determinantes del rendimiento académico durante los años mencionados, haciendo uso de la econometría espacial (Hernández, Fernández, & Baptista, 2014).

3.1.1 Econometría Espacial.

La econometría espacial es una rama de la econometría general que estudia los fenómenos económicos espaciales en los modelos de ciencia regional, la cual se basa en la premisa de que tanto la ubicación geográfica como la distancia son factores importantes para los estudios regionales (Acevedo & Velásquez, 2008).

Existen dos efectos espaciales, que son divididos en dos tipos: dependencia espacial o autocorrelación espacial y heterogeneidad espacial (Pérez, 2006).

3.1.1.1 *Autocorrelación Espacial.*

La autocorrelación espacial se origina como consecuencia de la existencia de una relación del valor de una variable en un lugar del espacio con su valor en otro u otros lugares del espacio, la cual puede ser positiva o negativa (Cliff & Ord, 1973).

Si la presencia de un fenómeno en una región lleva a que se expanda ese mismo fenómeno hacia el resto de regiones que la rodean, propiciando así un agrupamiento del mismo, se estará ante un caso de autocorrelación positiva. Al contrario, existirá autocorrelación espacial negativa cuando la presencia de un fenómeno no influya en la aparición de las regiones vecinas (Moreno & Vayá, 2004).

En el análisis de autocorrelación espacial se necesita un instrumento que fusione el hecho de la interdependencia y las relaciones multidireccionales de las variables, por eso se utiliza la matriz de pesos espaciales, la cual es definida con la letra **W** cuyos elementos representan la interdependencia existente entre las regiones (Getis, 2010).

En la Figura N.º3, se presenta los diversos tipos de contigüidad las cuales pueden ser de tres tipos principalmente; el primero es el caso torre, que cuenta solo las ubicaciones adyacentes; el tipo reina que considera un vecindario de ocho celdas mientras que el caso de alfil únicamente considera las diagonales de cada celda (Borrego, 2018).

Figura N.º3. Tipos de Contigüidad



Fuente: (Borrego, 2018)

3.1.1.1.1 Medida de Autocorrelación Espacial.

El análisis de autocorrelación espacial global es un estadístico para resumir toda la zona de estudio, por tanto, asume homogeneidad y esto permite que se produzca un valor para cualquier matriz de pesos espaciales. Es decir, contrasta la hipótesis de que una variable se encuentra distribuida de forma aleatoria en el espacio o si, por el contrario, existe una asociación significativa de valores similares o diferentes entre regiones vecinas (Celemín, 2009).

a. I de Moran Global.

El estadístico I de Moran global mide la autocorrelación espacial basada en las ubicaciones y los valores de las entidades simultáneamente. Asimismo, es una regla de decisión para evaluar en qué medida la distribución espacial cumple la hipótesis nula de que las áreas cercanas no se afectan entre sí, mientras que la hipótesis alternativa implica que las áreas se agrupan mediante algún tipo de autocorrelación espacial (Anselin, 2005).

La I de Moran viene dada por la siguiente expresión:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2},$$

donde n es el número de áreas que se considera y W_{ij} es una matriz de pesos definida anteriormente. En esta medida, la distancia que utiliza para crear la matriz M viene dada por $(z_i - \bar{z})(z_j - \bar{z})$ siendo z_i y z_j el valor de la variable Z del área i y j respectivamente, y siendo \bar{z} la media aritmética de todos los valores de las áreas que queremos estudiar.

Aunque la I de Moran es una de las medidas más antiguas, sigue siendo la medida más utilizada para determinar la autocorrelación espacial. Los resultados del análisis se interpretan dentro del contexto de la hipótesis nula que establece que la variable que se analiza está distribuida en forma aleatoria entre las entidades del área de estudio.

b. I de Moran Local.

El estadístico I_i local de Moran calcula, para cada unidad i , la existencia de un agrupamiento significativo de valores similares (altos o bajos) alrededor de dicha unidad geográfica del modo siguiente:

$$I_i = \frac{y_i}{\sum_{j \in J_i} W_{ij}} \sum_{j \in J_i} W_{ij} y_j,$$

siendo y_i y y_j los valores de la variable Y en las unidades espaciales i , j ; W_{ij} son los elementos de la matriz de pesos espaciales W y J_i es el conjunto de unidades relacionadas con i (Anselin, 1995).

Cuando esta expresión resulta un valor positivo indica la existencia de un agrupamiento espacial de valores similares alrededor de la unidad espacial i , en tanto que un valor negativo indica el agrupamiento de valores muy diferentes alrededor de i . Asimismo, también puede comprobarse que la suma de estadísticos locales I_i es el estadístico global I de Moran.

c. Mapa LISA.

El mapa LISA de “clúster” espacial es la representación gráfica del I_i local de Moran, en el cual los valores significativos de I_i se lleva a cabo a través de distintos colores, uno para cada cuadrante del diagrama de dispersión de Moran. De este modo, es posible saber alrededor de cuál de los puntos del diagrama se produce un agrupamiento significativo de datos similares o diferentes de una variable, así como el signo de la dependencia espacial (Anselin, 1995).

En resumen, la hipótesis de estacionariedad es la base sobre la que se asienta el análisis estadístico de la autocorrelación espacial. En el caso de existencia de dependencia espacial local, este supuesto se incumple y por eso el análisis global no sería válido por lo cual será necesario un indicador espacial local, el cual será mejor comprendido con su representación gráfica.

3.1.1.1.2 *Modelación Espacial.*

La literatura sobre Econometría espacial en la última década se caracteriza por el interés en el manejo de base de datos tipo panel, puesto que en la actualidad existe una mayor disponibilidad de bases de datos donde se combina una dimensión temporal con otra transversal (Paelinck, Mur, & Trívez, 2004). La dependencia espacial en modelos de regresión surge como consecuencia de autocorrelación en variables dependientes, independientes o en las perturbaciones. Estas variantes asociadas al tipo de dependencia espacial se traducen en diversas formas de incorporación a los modelos de correlación espacial a través de matrices de peso para su especificación (Pérez, 2006).

Se utiliza la técnica pooled data para hallar las estimaciones de los coeficientes de las variables explicativas, sin embargo, las críticas a este modelo es que omite posibles fuentes de heterogeneidad. Pues, no consideran particularidades de cada región que pueden ser específicas e invariantes en el tiempo ni recesiones que pueden causar diferencias entre periodos temporales (Montero, 2011).

Basado en la teoría de Hendry (2000), una vez detectada la importancia del espacio en el modelo panel data, se utilizan dos estrategias distintas para introducir los efectos espaciales en el modelo lineal general. La primera estrategia empleada es “de lo particular hacia lo general”, el cual se aplica mediante los estadísticos de Multiplicadores de Lagrange (LM) simples y robustos que permite determinar cuál es el modelo adecuado a usar, el modelo de retardo espacial (SAR) o el modelo de error espacial (SEM). Por otro lado, la segunda estrategia de especificación es “de lo general a lo particular”. El modelo general con los tres tipos de efectos tiene la siguiente expresión:

$$Y = \rho W_y + X\beta + W_x\theta + \varepsilon,$$

$$\varepsilon = \lambda W_\varepsilon + u,$$

donde W_y es el retardo espacial de la variable endógena, W_x representa los efectos de interacción exógena, W_ε , los de interacción en el error y ρ , λ , θ son los coeficientes de autocorrelación espacial vinculados a los respectivos mecanismos de interacción; W finalmente es la matriz de pesos, que describe la distribución sobre el espacio de las unidades de observación.

A partir del modelo general, se pueden obtener diferentes especificaciones. Se está en un modelo Spatial Durbin Model (SDM) cuando λ sea igual a 0. Asimismo, se establecen dos modelos en función del valor de diferentes parámetros; así, cuando $\rho = 0$ y $\theta = 0$ se está frente a un modelo Spatial Error Model (SEM); y por último cuando $\theta = 0$ y $\lambda = 0$ se está ante un modelo Spatial Lag Model (SAR), cuyas expresiones finales son:

a. Modelo SDM.

LeSage & Pace (2009) proponen al modelo SDM como la mejor alternativa de especificación general por dos razones. Primero, porque la omisión de un término relevante en el error sólo genera pérdida de eficiencia; sin embargo los costos de ignorar efectos espaciales en la variable dependiente y/o la omisión de variables explicativas relevantes generan estimadores sesgados e inconsistentes de los coeficientes.

$$Y = \rho W_y + X\beta + W_x\theta + u$$

b. Modelo SAR.

El Modelo Espacial Autorregresivo incorpora el rezago espacial de la variable dependiente, el cual permite describir la dependencia espacial cuando los valores que adopta una variable en una región están relacionados con las observaciones vecinas (Anselin, 2005).

$$Y = \rho W_y + X\beta + u$$

c. Modelo SEM.

El modelo de error espacial supone que la dependencia espacial hallada en la variable dependiente es resultado de la distribución geográfica de las variables explicativas y de la autocorrelación del término del error.

$$Y = X\beta + \lambda W\varepsilon + u,$$

Asimismo, Anselin (2005) afirma también que ignorar la autocorrelación espacial del error genera que los estimadores de la regresión son ineficientes. Sin embargo, dado que este modelo no tiene rezagos espaciales de la variable dependiente, los parámetros β estimados se interpretan como derivadas parciales.

3.1.1.2 Heterogeneidad Espacial.

La heterogeneidad espacial es un efecto de la econometría espacial que está relacionado con la diferenciación espacial de las unidades geográficas, es decir que los coeficientes del modelo varían según la ubicación. Esto implicará que los modelos espaciales y los parámetros variarán de acuerdo con la localización geográfica (Chasco, 2004).

De este modo, la heterogeneidad espacial surge cuando se utilizan datos de unidades espaciales muy distintas para explicar un mismo fenómeno. En ese caso, se manifiesta problemas de heterocedasticidad o inestabilidad estructural, que pueden ser resueltos mediante las técnicas econométricas normales (Moreno & Vayá, 2004).

3.2 Población y Muestra.

3.2.1 Población Objetivo.

En la investigación realizada es importante definir la población de estudio, que es definido como el conjunto de todos los elementos llamados también unidades de análisis que pertenecen al ámbito espacial donde se lleva a cabo el trabajo de investigación para la generalización de las conclusiones (Carrasco, 2005).

En la presente investigación, la población objetivo está conformada por las regiones del país en los cuales se rinde el examen de la Evaluación Censal para el periodo comprendido entre los años 2012 y 2016.

3.2.2 Método de muestreo.

En el presente trabajo se utilizará el método de muestreo no probabilístico por conveniencia. En ese sentido, la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación. Siendo así, la muestra de la presente investigación estará conformada por las 26 regiones del Perú, desagregando la región Lima en Lima Metropolitana, Lima Provincias y El Callao, durante cinco años que comprende desde el 2012 a 2016 (Hernández, Fernández, & Baptista, 2014).

3.2.3 Tamaño de la muestra.

Por ser una investigación con un análisis de panel data, para identificar los determinantes del rendimiento académico a nivel general se ha decidido trabajar con el periodo comprendido entre los años 2012 a 2016 de forma anual, a nivel regional. Ello nos da 26 observaciones por cada 5 años por cada una de las variables. Asimismo, para identificar los determinantes del rendimiento académico según ámbito geográfico se trabajará con el periodo comprendido entre los años 2012 a 2016 para las regiones de la Costa (12) y Sierra (9); mientras que, para la Selva (5) por un tema de insuficiente número de observaciones se trabajará con el periodo comprendido entre los años 2009 a 2016, lo cual nos da 60, 45 y 40 observaciones respectivamente.

3.3 Método de Recolección de Datos.

Para la recolección de datos, se utiliza fuentes secundarias, es decir datos de una institución pertinente y con el prestigio respectivo que asegura la calidad de estos. En este estudio, se ha utilizado datos estimados por el Ministerio de Educación, por su confiabilidad y consistencia.

El puntaje promedio de los Resultados ECE en Comprensión Lectora, fue obtenido de la base de datos del Sistema de Consulta de Resultados de Evaluaciones (SICRECE, 2018) del MINEDU. Para las variables de servicios básicos, mobiliario escolar acceso a internet; así como también educación inicial, tamaño de clase y profesores titulados se utilizaron las Estadísticas de la Calidad Educativa (ESCALE, 2018) del MINEDU. Por último, para el indicador de Necesidades Básicas Insatisfechas (NBI) se usó la base de datos del Instituto Nacional de Estadística e Informática (INEI, 2018).

Con esta información se elabora la base de datos del tipo *panel data*, que brinda al investigador un gran número de datos, aumentando los grados de libertad y reduciendo la colinealidad entre las variables explicativas, lo que hace que la estimación econométrica sea más eficiente (Hsiao, 2003).

3.4 Método de Análisis de Datos.

En primer lugar, se procederá a realizar el análisis exploratorio de datos espaciales (AEDE) que consiste en crear una matriz con las propiedades de los datos y su ubicación. Utilizando esta matriz, se analizará si existe dependencia espacial, es decir, si los valores de la variable en una región están relacionados con los valores de las regiones cercanas. Para realizar este análisis se utilizará indicadores que medirán la dependencia espacial y contrasten la hipótesis con el índice I de Moran.

Después de hacer el análisis exploratorio de datos, se llevara a cabo el análisis confirmatorio de los datos espaciales, en el que se formula modelos de regresión y se realiza la estimación de parámetros muestrales, en la cual se determinará si el modelo presenta dependencia espacial o no. El modelo de regresión espacial que se tomarán para el análisis confirmatorio será el modelo SDM y el modelo SEM para panel data por la naturaleza de nuestros datos que combina información sobre el tiempo y el espacio.

Finalmente, se utilizará la función de producción basada en Glewwe & Kremer (2006) donde el rendimiento escolar depende de factores socioeconómicos, de la calidad del docente, de las condiciones de la escuela y de las características del alumno.

$$\text{Log}(A_{it}) = \beta_0 + \beta_1 * SB_{it} + \beta_2 * ME_{it} + \beta_3 * IN_{it} + \beta_4 * EI_{it} + \beta_5 * TC_{it} + \beta_6 * PT_{it} + \beta_7 * NBI_{it} + \mu_{it}$$

Donde A_{it} es el puntaje promedio de los Resultados ECE en Comprensión Lectora, SB_{it} es el porcentaje de escuelas que tienen acceso a los tres servicios básicos (electricidad, agua y desagüe), ME_{it} es el porcentaje de escuelas que cuentan con suficiente carpetas, sillas, etc. ; IN_{it} es el porcentaje de escuelas que cuentan con acceso a Internet en primaria, EI_{it} es el porcentaje de ingresantes a primaria con educación inicial, TC_{it} es el número de alumnos por clase en nivel primaria, PT_{it} es el porcentaje de profesores titulados en primaria, NBI_{it} es el porcentaje de necesidad básica insatisfecha en cada región, finalmente μ_{it} es el error que representa los factores inobservables que cambia con el tiempo.

3.5 Técnicas para el Procedimiento y Análisis de la Información.

El desarrollo del trabajo se elaborará a través del programa Stata (Software Statistics and Data Science) que es un /software/ de estadística completa e integrada que permite gestionar, analizar y describir los datos por medio de métodos estadísticos y gráficos. La facilidad del manejo de este programa informático permitirá tratar los datos panel desde una dimensión temporal para la obtención del modelo óptimo del trabajo de investigación.

Para la matriz de pesos, el estadístico I de Moran tanto global como local, los mapas LISA y los mapas por clúster se realizará a través del Software Geo estadístico GeoDa, que permite describir los datos mediante el uso de métodos gráficos. Del mismo modo, facilita el análisis de asociación espacial de los datos a través de indicadores de dependencia espacial que identifica patrones de agrupamiento o dispersión de los datos (Anselin, 2005).

CAPITULO IV: ANÁLISIS DE LA INTERPRETACIÓN DE LAS PRUEBAS ESTADÍSTICAS, FÓRMULAS, RESULTADOS

Este capítulo se dividirá en dos apartados, en el primero se realizará el Análisis Estadístico de Datos Espaciales en el cual se verá el estadístico I de Moran Global y local y en el segundo apartado se establecerá qué modelo de regresión espacial es el más adecuado para el análisis de resultados tanto a nivel regional como según el ámbito geográfico, pues no se pueden aplicar políticas públicas iguales a realidades diferentes.

4.1 Modelo de regresión general.

4.1.1 Análisis Exploratorio de Datos Espaciales.

En este apartado, presentamos algunos métodos básicos de AEDE, destinados no solo a la representación de fenómenos espaciales, sino también a una previa presentación de la realidad de las regiones del Perú.

A continuación, analizaremos los determinantes del rendimiento académico a nivel general incluidos en nuestro modelo a través del diagrama de dispersión de Moran, el mapa LISA y el mapa por clúster.

4.1.1.1 Rendimiento Académico.

En el *Anexo N°7.2*, se observa que el índice de Moran obtenido para esta variable expresa autocorrelación espacial medianamente significativa. La situación que refleja la prueba de autocorrelación para esta variable implica un rechazo de la hipótesis nula de la existencia de aleatoriedad espacial, por lo cual se espera que el puntaje promedio del rendimiento académico para diferentes regiones sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

En el mapa LISA se puede distinguir la categoría High-High (color rojo) que se refiere a las fracciones censales que tienen alto puntaje promedio del rendimiento académico y que, a su vez, se encuentran rodeadas de otras regiones que también registran este tipo de valor y en el otro extremo, la categoría Low - Low (color azul) hace alusión a las regiones que tienen bajo puntaje promedio del rendimiento académico y se encuentran rodeadas de otras regiones que en

promedio también registran este tipo de valor, con lo que se hace visible su agrupamiento o autocorrelación espacial positiva en el área de la Costa del Perú.

Asimismo, en el mapa por clúster se observa qué regiones son las que tienen mejor rendimiento académico (color verde), cuáles se encuentran en el promedio (color azul) y qué regiones tienen menor rendimiento académico (color celeste). Como se puede observar del gráfico, en el año 2016, las regiones que tienen mayor rendimiento académico son Lima Provincia, Moquegua, Arequipa y Tacna mientras que las que tienen menor rendimiento académico son Loreto y Ucayali.

4.1.1.2 Servicios Básicos.

En el *Anexo N° 7.3*, el I de Moran obtenido para esta variable expresa autocorrelación espacial medianamente significativa. La situación que refleja la prueba de autocorrelación para esta variable implica un rechazo de la hipótesis nula de la existencia de aleatoriedad espacial, por lo cual se espera que el porcentaje de los servicios básicos para diferentes regiones sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

En el mapa LISA se puede distinguir la categoría High-High que se refiere a las regiones que tienen mayor porcentaje de escuelas públicas con accesibilidad a los tres servicios básicos que, a su vez, se encuentran rodeadas de otras regiones que también registran este tipo de valor y en el otro extremo, la categoría Low-Low que hace alusión a las regiones que tienen menor porcentaje y se encuentran rodeadas de otras regiones que en promedio también registran el mismo valor, con lo que se hace visible su agrupamiento o autocorrelación espacial positiva en el área de la Costa del Perú.

Asimismo, en el mapa por clúster se observa que para el año 2016, las regiones que tienen mayor porcentaje de escuelas públicas que cuentan con 3 servicios básicos son Lima (incluyendo sus 3 regiones, Tumbes, Ancash, Ica, Moquegua, Arequipa y Tacna mientras que las que tienen menor porcentaje de escuelas públicas que cuentan con tres servicios básicos son Loreto, Ucayali, y Madre de Dios.

De la teoría y del análisis descriptivo se puede inferir que el aumento de escuelas públicas que cuentan con acceso a los tres servicios básicos como mínimo se traduce en mejores rendimientos académicos de los estudiantes. Así mismo se puede observar que las regiones de la Costa tienen mayor acceso a estos tres servicios básicos que la región Sierra y en menor porcentaje las regiones pertenecientes a la Selva (Elias & Rey, 2011).

4.1.1.3 *Mobiliario Escolar.*

En el *Anexo N° 7.4*, el I de Moran obtenido para la variable expresa, para los años 2012 a 2014, autocorrelación negativa. Sin embargo, para el año 2016 se observa autocorrelación positiva de 0.3. Para este último año de estudio se rechaza la hipótesis nula de existencia de aleatoriedad espacial, por lo cual se espera que esta variable, para diferentes regiones sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

En el mapa LISA se puede distinguir la categoría High-High que se refiere a las regiones que tienen mayor porcentaje de escuelas que cuentan con suficientes carpetas se encuentran rodeadas de otras regiones que también registran este tipo de valor. Mientras que la categoría Low – Low hace alusión a las regiones que tienen menor porcentaje de escuelas que cuentan con suficientes carpetas, con lo que se hace visible su agrupamiento positivo en el área de la selva del Perú.

Asimismo, en el mapa por clúster se observa para el año 2016, que las regiones que tienen menor porcentaje de escuelas con suficientes carpetas son las regiones de la Selva, de las cuales sobresalen Ucayali y Madre de Dios. Del análisis descriptivo se puede inferir que la relación del porcentaje de escuelas con suficientes carpetas con del rendimiento académico presenta una relación directa por lo que se espera un coeficiente positivo en esta variable.

4.1.1.4 Acceso a Internet.

En el *Anexo N° 7.5*, el I de Moran obtenido para la variable expresa autocorrelación espacial medianamente significativa. Se rechaza la hipótesis nula de la existencia de aleatoriedad espacial, por lo cual se espera que el porcentaje de escuelas que cuentan con acceso a internet para diferentes regiones sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

En el mapa LISA se puede distinguir la categoría High-High y la categoría Low – Low donde cada una se encuentra rodeada de otras regiones que en promedio también registran el mismo tipo de valor respectivamente. Asimismo, en el mapa por clúster se puede observar que en el año 2016, las regiones que tienen menor porcentaje de escuelas con acceso a internet son Loreto, Amazonas, San Martín y Madre de Dios. Por el contrario, las regiones como Tacna, Arequipa, Ica Lambayeque, Lima Provincia y Tumbes cuentan con un mayor porcentaje de escuelas con acceso a internet. Con estos resultados se infiere que en la región Selva hay menos escuelas con acceso a internet que en la Costa.

De este análisis y de la teoría vista con antelación, se puede inferir que la relación del porcentaje de escuelas que tienen acceso a internet con las variables que mide el rendimiento académico presenta una relación directa por lo que se espera un coeficiente positivo en esta variable.

4.1.1.5 Educación Inicial.

En el *Anexo N° 7.6*, se observa que el I de Moran obtenido para la variable expresa autocorrelación espacial poco significativa, asimismo se observa que se acepta la hipótesis nula de existencia de aleatoriedad espacial por lo que las ubicaciones de las regiones no tienen relación con el porcentaje de ingresantes a primaria con 3 años de educación inicial.

En el mapa LISA, se observa que con el pasar de los años esta variable se va concentrando en la categoría Low - Low, la cual hace alusión a las regiones que tienen menor porcentaje de niños con 3 años de educación inicial. Asimismo, en el mapa por clúster, se observa que las regiones que presentan este problema son Ucayali, Madre de Dios, Cuzco y Puno.

De estos análisis se puede inferir que la relación del porcentaje de niños con 3 años de educación inicial es directa con el rendimiento académico, por lo que en el análisis confirmatorio se espera un coeficiente positivo en esta variable.

4.1.1.6 *Tamaño de clase.*

En el *Anexo N° 7.7*, el I de Moran obtenido para la variable expresa autocorrelación espacial medianamente significativa. Se rechaza la hipótesis nula de la existencia de aleatoriedad espacial, por lo cual se espera que el tamaño de clase para diferentes regiones sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

En el mapa LISA se puede distinguir la categoría High-High y la categoría Low – Low donde cada una se encuentra rodeada de otras regiones que en promedio también registran el mismo valor respectivamente. Asimismo, en el mapa por clúster se puede observar que, la región Loreto, San Martín Ucayali y Madre de Dios tienen más número de alumnos por clase que la región Lima Provincia, Arequipa, Tumbes entre otras. De esta manera se afirma que el tamaño de clase en la Selva es mayor que en la Costa.

Asimismo, se observa que el tamaño de clase tiene una relación inversa con el rendimiento académico, por lo que se espera un coeficiente negativo en esta variable.

4.1.1.7 *Profesores Titulados.*

En el *Anexo N° 7.8*, se observa que el I de Moran obtenido para la variable expresa una autocorrelación espacial medianamente significativas, asimismo se observa que se acepta la hipótesis nula de existencia de aleatoriedad espacial. Por lo que el porcentaje de profesores titulados no se ve influido por las ubicaciones de las regiones.

Sin embargo en el mapa por clúster, se observa que las regiones que cuentan con mayores porcentajes de profesores titulados son Ucayali, Madre de Dios, Cuzco, Puno. Este análisis descriptivo es un poco ambiguo, ya que se esperaba que estas regiones sean las que menor porcentaje de profesores titulados

haya sin embargo es al contrario, por lo cual se esperara a hacer un análisis confirmatorio para su interpretación.

4.1.1.8 *Necesidades Básicas Insatisfechas.*

En el *Anexo N° 7.9*, el índice I de Moran es medianamente significativa, asimismo se puede observar que se rechaza la hipótesis nula de existencia de aleatoriedad espacial, por lo cual se espera que el porcentaje de la población con al menos una Necesidad Básica Insatisfecha (NBI) para diferentes regiones sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

Con el mapa LISA se afirma que hay una autocorrelación espacial positiva en las regiones de la Selva como Loreto, Amazonas, San Martín, Huánuco y Ucayali que a su vez representan las áreas que tienen mayor población con al menos una NBI.

De la teoría y del análisis descriptivo se puede inferir que la relación de esta variable con las variables del rendimiento académico y acceso a los tres servicios básicos presenta una relación inversa por lo que se espera un coeficiente negativo en esta variable.

4.1.2 Modelación Espacial del modelo general.

En este apartado se establecerá qué modelo de regresión es el más adecuado para el análisis de los resultados.

4.1.2.1 Análisis del modelo general.

En primer lugar, se realiza la estimación del modelo para analizar qué determinantes del rendimiento académico en Comprensión Lectora tienen estadísticos significativos. Asimismo, se verifica a través del test de Durbin-Wu-Hausman en el que el modelo no presenta problema de endogeneidad (*Anexo N.º 7.10*). Basándose en los resultados de la Tabla N.º2, se piensa que hay un problema de colinealidad, pues la mayoría de las variables no tienen significancia en el modelo y además algunos signos de los estimadores son distintos a los que se espera según la teoría. Sin embargo, a través del factor inflacionario de la varianza se constata que el problema de multicolinealidad es moderada. Por ello, se procede a elegir qué modelo de regresión se debe utilizar para realizar el análisis de los resultados del modelo general.

Tabla N.º2. Regresión del modelo general - Pooled Data

A2	Coef.	Z	$p > z $
<i>SB</i>	-0,0006463	-0,99	0,324
<i>ME</i>	-0,0007257	-0,89	0,372
<i>IN</i>	0,0021146	3,87	0,000
<i>EI</i>	0,0016272	5,50	0,000
<i>TC</i>	-0,0125421	-4,78	0,000
<i>PT</i>	-0,001126	-1,47	0,140
<i>NBI</i>	-0,0010478	-1,37	0,170
<i>Cons.</i>	6,604601	53,87	0,000

$N = 130$ $R^2 = 0,54$ $VIF = 3,98$ $Wald\ chi^2 = 276,88$ $Prob > \chi^2 = 0,00$ $\rho = 0,59$

Elaboración Propia

4.1.2.1.1 Elección del modelo óptimo

a. De lo particular hacia lo general

Para elegir el modelo óptimo haciendo uso de esta estrategia, es de importancia estimar los Multiplicadores de Lagrange. Estos test nos van a dar información sobre la existencia o no de autocorrelación espacial en el modelo y sobre la forma de incorporación de la misma (Anselin, 2005).

Como se observa en la Tabla N°3, los p-value de los modelos de LM-error como el de LM-lag son menores a 0.05, por lo tanto, se tienen efectos espaciales significativos y se necesitara usar modelos espaciales en lugar del modelo MCO. Como ambas estadísticas de la prueba LM rechazan la hipótesis nula de ausencia de dependencia espacial, se considera las formas robustas de las estadísticas de prueba. En este caso, la decisión es simple, ya que se estimará el modelo de regresión espacial error SEM, ya que esta es más significativa que la de retardo.

A modo de resumen, en el *Anexo N.º7.10.1* se muestra cómo según esta estrategia de especificación se elige el modelo SEM para el modelo general (Anselin, 2005).

Tabla N°3. Dependencia espacial en regresión MCO

<i>Test</i>	<i>Statics</i>	<i>Df</i>	<i>p – value</i>
<i>Spatial error :</i>			
<i>Lagrange Multiplier</i>	13,509	1	0,000
<i>Robust Lagrange Multiplier</i>	7,091	1	0,008
<i>Spatial lag :</i>			
<i>Lagrange Multiplier</i>	7,386	1	0,007
<i>Robust Lagrange Multiplier</i>	0,967	1	0,325

Elaboración Propia

b. De lo general hacia lo particular

Haciendo uso de esta estrategia, en el *Anexo 7.10.2* se muestran los resultados del modelo general SDM y se compara con el modelo SEM. Asimismo, se aplica el test de Hausman con el fin de determinar qué modelo es más adecuado, si el de efectos aleatorios o el de efectos fijos.

En el modelo SDM, el resultado de la prueba de Hausman determina un χ^2 de 13.62 y una $\text{prob} > \chi^2$ igual a 0.0000 (menor de 0.05), lo que indica que se debe rechazar la hipótesis nula, por lo tanto, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas.

Por otro lado, en el modelo SEM, el resultado de la prueba de Hausman determina un χ^2 de 23.4 y una $\text{prob} > \chi^2$ igual a 0.0015 (menor de 0.05), lo que indica que se debe rechazar la hipótesis nula, por lo tanto, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas.

- **MODELO SDM vs SEM**

En la Tabla N.º4, se comparan los resultados de los coeficientes de los modelos SDM y SEM con efectos fijos, con sus respectivas probabilidades. Para el modelo *Sdmfe*, el rendimiento académico de la región está explicado por el rendimiento académico de sus vecinos en un 69%, mientras que para el modelo *Semfe*, el rendimiento académico de la región se explica por variables que no están en el modelo de los vecinos más sí en un 75% en los términos error. Con base al criterio BIC, se elige el modelo SEM, en la que se observa que tanto la variable de mobiliario escolar, acceso a internet, profesores titulados y el NBI no muestran una relación significativa con el rendimiento académico, por lo cual se procede a sacarlos del modelo.

Tabla N°4. Resumen de Resultados de Coeficientes

<i>Variable</i>		<i>Sdmfe</i>	<i>Semfe</i>
Main	SB	0,00071365	0,0010754*
	ME	-0,00019548	-0,00055674
	IN	0,00003143	0,00054644
	EI	0,00060836	0,00141282***
	TC	-0,01422226***	-0,01488939***
	PT	0,00037624	0,00011065
	NBI	0,00037757	0,00036418
Rho		0,68967474***	
Lambda		0,7513989***	
R-sq		0,37	0,38
BIC		-672,03	-688,08

$N = 130$ * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Elaboración Propia

4.1.2.1.2 Modelo General Corregido.

Como ya se ha seleccionado el modelo SEM de efectos fijos como el óptimo y se ha determinado que se tiene que eliminar del modelo un conjunto de variables estadísticamente no significativos, el modelo a estimar es el siguiente:

$$\text{Log}(A_{it}) = \beta_0 + \beta_1 * SB_{it} + \beta_2 * EI_{it} + \beta_3 * TC_{it} + \lambda W_{\varepsilon} + \mu_{it}$$

Tabla N° 5. Resultados finales de la estimación econométrica

<i>A2</i>	<i>Coef.</i>	<i>Z</i>	<i>p > /z/</i>
<i>SB</i>	0,0011266	1,92	0,055
<i>EI</i>	0,0012625	3,15	0,002
<i>TC</i>	-0,0131852	-6,28	0,000
<i>Cons.</i>	6,4324		
<i>Lambda</i>	0,7926		

$$N = 130 \quad R^2 = 0,35$$

Elaboración propia

Basándose en los resultados observados en la Tabla N.º5, el modelo general final queda de la siguiente forma:

$$\text{Log}(A_{it}) = 6.43 + 0.0011 * SB_{it} + 0.0013 * EI_{it} - 0.013 * TC_{it} + 0.80 W_{\varepsilon} + \mu_{it}$$

Dado que este modelo no tiene rezagos espaciales de la variable dependiente, los parámetros estimados se interpretan como derivadas parciales. Además, al tener una forma funcional semi - logarítmica, se obtienen semi - elasticidades, por lo que un cambio en uno de los indicadores educativos ocasionará una variación porcentual en el rendimiento académico.

A nivel general, los determinantes del rendimiento académico en Comprensión Lectora son los servicios básicos, la educación inicial y el tamaño de clase. Ante un incremento en el porcentaje de escuelas públicas que cuentan con servicios básicos; el rendimiento aumenta en un 0,11%. Asimismo, ante un incremento en el porcentaje de alumnos matriculados que cursaron al menos tres años de educación inicial; el rendimiento aumenta en un 0,13%. De igual manera, ante el incremento de un alumno por clase; el rendimiento disminuye en un 1,3%. Por último, en cuanto a la dependencia espacial se concluye a nivel general que, el rendimiento académico en Comprensión Lectora se explica por variables que no están en el modelo de las regiones vecinas mas si en un 80% en los términos error.

4.1.3 Análisis de los Resultados a nivel general.

En primer lugar, es preciso señalar que para el presente trabajo de investigación se toma la variable de rendimiento académico en Comprensión Lectora como variable proxy del rendimiento académico, lo cual genera una limitación para la interpretación de los resultados.

Los resultados encontrados sugieren que el rendimiento académico de una región está relacionado en un 80% con el rendimiento de sus regiones colindantes, ya que se encuentran en ubicaciones similares y comparten las mismas características. (Elías & Rey, 2011). En el análisis exploratorio de la variable dependiente, las regiones se agrupan de manera que las regiones de la Costa tienen un mayor rendimiento académico que la Sierra y Selva.

En base a las tendencias que se observa en el *Anexo 7.11*, estas diferencias de los resultados del rendimiento académico en Comprensión Lectora a nivel regional se deben a que en la región Costa la mayoría de las escuelas cuentan con acceso a servicios básicos, además la mayoría de alumnos ingresantes a primaria cuentan con 3 años de educación inicial y las escuelas tienen un menor promedio de tamaño de clase; que en conjunto hacen que el rendimiento académico de la Costa se encuentre por encima de la región Sierra y Selva.

Por otro lado, se resalta que la ecuación de regresión final sirve en caso se quiera aplicar modelos predictivos del rendimiento académico en el Perú. A partir de los resultados obtenidos del modelo, se afirma que los determinantes que impactan sobre el rendimiento académico en Comprensión Lectora a nivel general se deben al acceso de los servicios básicos en la escuela (Niu, Jia, Wang, He, Hu, & Liu, 2013), a la educación inicial (Currie & Thomas, 2000; Berlisnki, Galiani & Gertler, 2006) y al tamaño de clase (Mosteller, 1995).

En ese sentido, diversos autores (Niu, Jia, Wang, He, Hu & Liu, 2013; UNICEF, 2017) señalan que las escuelas aumentan su tasa de asistencia escolar al contar con los servicios básicos, lo cual afecta directamente sobre el rendimiento académico de los alumnos. Asimismo, diversos autores, también señalan que la educación inicial es esencial para el rendimiento académico, (Currie & Thomas, 2000; Berlisnki, Galiani & Gertler, 2006) pues a largo plazo origina un círculo virtuoso de desarrollo educativo favoreciendo la igualdad de oportunidades. Por último (Mosteller, 1995) afirma que el tamaño de clases sí influye en el rendimiento académico de los alumnos pues este resulta ser mayor cuando las clases son más pequeñas.

4.2 Modelo de regresión según ámbito geográfico.

El sistema educativo todavía no ha logrado reducir las brechas educativas entre los estudiantes de diferentes regiones del país, caracterizadas especialmente por sus diferencias socioeconómicas y de ubicación geográfica (MINEDU, 2016b). Por ello, se elaboraron modelos específicos del rendimiento académico en Comprensión Lectora diferenciados según región geográfica: Costa, Sierra y Selva.

4.2.1 Análisis Exploratorio de Datos Espaciales.

En este apartado, presentamos algunos métodos básicos de AEDE, para realizar una previa presentación de la realidad del rendimiento académico en las regiones de la Costa, Sierra y Selva del Perú.

A continuación, analizaremos al rendimiento académico a través del diagrama de dispersión de Moran, el mapa LISA y el mapa por clúster.

4.2.1.1 Rendimiento Académico de la Costa.

En el *Anexo N°7.12*, se observa que el índice de Moran obtenido para esta variable expresa autocorrelación espacial significativa. La situación que refleja la prueba de autocorrelación para esta variable implica un rechazo de la hipótesis nula de la existencia de aleatoriedad espacial, por lo cual se espera que el puntaje promedio del rendimiento académico para la región Costa sea más similar en las regiones vecinas que en aquellas separadas por grandes distancias.

Asimismo, el mapa por clúster brinda también una información importante de la variable, en la cual se ve qué regiones son las que tienen mejor rendimiento académico (color verde), cuáles se encuentran en el promedio (color azul) y qué regiones tienen menor rendimiento académico (color celeste). Como se puede observar del gráfico, en el año 2016, las regiones que tienen mayor rendimiento académico en la Costa son Moquegua y Tacna mientras que las que tienen menor rendimiento académico son Tumbes, Piura, La Libertad y Ancash.

4.2.1.2 Rendimiento Académico de la Sierra.

En el *Anexo N°7.13*, el índice de Moran obtenido para la variable expresa una autocorrelación negativa. Asimismo se observa que se acepta la hipótesis nula de existencia de aleatoriedad espacial. Por lo que el rendimiento académico de los estudiantes en la región Sierra no se ven influenciados por las ubicaciones de las regiones.

Por otro lado, el mapa por clúster brinda también una información importante de la variable, en la cual se ve que para el año 2016, las regiones de la Sierra que tienen mejor rendimiento académico son Junín, Ayacucho, Cusco y Puno mientras que las que tienen menor rendimiento académico son Cajamarca, Huánuco y Apurímac.

4.2.1.3 Rendimiento Académico de la Selva.

En el *Anexo N.° 7.14*, el índice de Moran obtenido para la variable expresa una autocorrelación medianamente significativa. Asimismo, se observa que se acepta la hipótesis nula de existencia de aleatoriedad espacial. Por lo que el rendimiento académico de los estudiantes en la región Selva no se ven influenciados por las ubicaciones de las regiones.

Por otro lado, el mapa por clúster brinda también una información importante de la variable, en la cual se ve que para el año 2016, las regiones de la Selva que tienen mejor rendimiento académico son Amazonas, San Martín y Madre de Dios; mientras que la que tiene menor rendimiento académico es Loreto.

4.2.2 Modelación Espacial según ámbito geográfico.

4.2.2.1 Modelación Espacial para la región Costa.

En este apartado se establecerá qué modelo de regresión espacial para la región Costa es el más adecuado para el análisis de los resultados.

4.2.2.1.1 Análisis del modelo para la región Costa.

En primer lugar, se realiza la estimación del modelo para analizar qué determinantes del rendimiento académico en Comprensión Lectora resultan estadísticamente significativos para la región Costa. Asimismo, se verifica a través del test de Durbin-Wu-Hausman en el que el modelo no presenta problema de endogeneidad (*Anexo N.º 7.15*). Basándose en los resultados de la Tabla N.º 6 se piensa que hay un problema de colinealidad, pues la mayoría de las variables no tienen significancia en el modelo y además algunos signos de los estimadores obtenidos con las pruebas individuales son distintos a los que se espera según la teoría. Sin embargo, a través del factor inflacionario de la varianza se constata que el problema de multicolinealidad es moderada. Por ello, se procede a elegir qué modelo de regresión se debe utilizar para realizar el análisis de los resultados del modelo para la región Costa.

Tabla N°6. Regresión del modelo para la región Costa - Pooled Data

A2	Coef.	Z	$p > z $
<i>SB</i>	-0,0002749	-0,27	0,786
<i>ME</i>	-0,0001274	-0,10	0,924
<i>IN</i>	0,0012858	1,70	0,090
<i>EI</i>	0,0014397	3,38	0,001
<i>TC</i>	-0,0067223	-1,36	0,173
<i>PT</i>	-0,0005076	-0,57	0,571
<i>NBI</i>	-0,0022504	-1,42	0,156
<i>Cons.</i>	6,446572	35,88	0,000

$N = 60$ $R^2 = 0,38$ $VIF=4,51$ $Wald\ chi^2 = 83,16$ $Prob > \chi^2 = 0,00$ $\rho = 0,77$

Elaboración Propia

4.2.2.1.1.1 Elección del modelo óptimo para la región Costa.

a. De lo particular hacia lo general

Para elegir el modelo óptimo haciendo uso de esta estrategia, es de importancia estimar los Multiplicadores de Lagrange. Como se observa en la Tabla N.º7, los p-value de los modelos de LM-error como el de LM-lag son menores a 0.05, por lo tanto, se tiene efectos espaciales significativos y se necesitara usar modelos espaciales en lugar del modelo MCO. Como ambas estadísticas de la prueba LM rechazan la hipótesis nula de ausencia de dependencia espacial, se considera las formas robustas de las estadísticas de prueba. En este caso, la decisión es simple, ya que se estimará el modelo de regresión espacial error SEM, ya que esta es más significativa que la de retardo.

Tabla N°7. Dependencia espacial en regresión MCO

<i>Test</i>	<i>Statics</i>	<i>Df</i>	<i>p – value</i>
<i>Spatial error :</i>			
<i>Lagrange Multiplier</i>	14,955	1	0,000
<i>Robust Lagrange Multiplier</i>	16,812	1	0,000
<i>Spatial lag :</i>			
<i>Lagrange Multiplier</i>	4,694	1	0,030
<i>Robust Lagrange Multiplier</i>	6,551	1	0,010

Elaboración Propia

b. De lo general hacia lo particular

Haciendo uso de esta estrategia, en el *Anexo 7.15.1* se estima el modelo general SDM y se compara con el modelo SEM. Asimismo, se aplica el test 15 de Hausman con el fin de determinar qué modelo es más adecuado, si el de efectos aleatorios o el de efectos fijos.

En el modelo SDM, el resultado de la prueba de Hausman determina un χ^2 de 82.14 y una $\text{prob} > \chi^2$ igual a 0.0000 (menor de 0.05), lo que indica que se debe rechazar la hipótesis nula, por lo tanto, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas.

Por otro lado, en el modelo SEM, el resultado de la prueba de Hausman determina un χ^2 de 15.47 y una $\text{prob} > \chi^2$ igual a 0.0304 (menor de 0.05), lo que indica que se debe rechazar la hipótesis nula, por lo tanto, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas.

• **MODELO SDM vs SEM**

En la Tabla N.º8, se comparan los resultados de los coeficientes de los modelos SDM y SEM, con sus respectivas probabilidades. Para el modelo *Sdmfe*, el rendimiento académico de la Costa está explicado por el rendimiento académico de sus colindantes en un 44%, mientras que para el modelo *Semfe*, el rendimiento académico de la Costa se explica por variables que no están en el modelo de los vecinos más si en un 53% en los términos error.

En base al criterio BIC, se elige el modelo SEM, en la que se observa que tanto la variable de servicios básicos, mobiliario escolar, tamaño de clase, profesores titulados y el NBI no muestran una relación significativa con el rendimiento académico, por lo cual se procede a sacarlos del modelo.

Tabla N°8. Resumen de Resultados de Coeficientes

<i>Variable</i>		<i>Sdmfe</i>	<i>Semfe</i>
Main	SB	0,00036507	0,00127138
	ME	-0,00011495	0,00004215
	IN	-0,00033123	0,00131406*
	EI	0,00028316	0,00083465*
	TC	-0,00605007	-0,01038552
	PT	-0,00122333	0,00014756
	NBI	0,00035585	-0,00057524
Rho		0,44447815***	
Lambda		0,52694235***	
R-sq		0,005	0,31
BIC		-295,32	-295,36

$N = 60$ * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Elaboración Propia

4.2.2.1.1.2 *Modelo Corregido para la región Costa.*

Como ya se ha seleccionado el modelo SEM de efectos fijos como el óptimo y se ha determinado que se tiene que eliminar del modelo un conjunto de variables estadísticamente no significativos, el modelo a estimar es el siguiente:

$$\text{Log}(A_{it}) = \beta_0 + \beta_1 * IN_{it} + \beta_2 * EI_{it} + \lambda W_{\varepsilon} + \mu_{it}$$

Tabla N° 9. Resultados finales de la estimación econométrica

A2	Coef.	Z	p > /z/
IN	0,0014165	2,61	0,009
EI	0,0013363	2,87	0,004
Cons.	6,2247		
Lambda	0,44		

$$N = 60 \quad R^2 = 0,15$$

Elaboración propia

Basándose en los resultados observados en la Tabla N.º9, el modelo para la región Costa queda de la siguiente forma:

$$\text{Log}(A_{it}) = 6.22 + 0.0014 * IN_{it} + 0.0013 * EI_{it} + 0.44 W_{\varepsilon} + \mu_{it}$$

En la región Costa, los determinantes del rendimiento académico en Comprensión Lectora son el acceso a internet y la educación inicial. Ante un incremento en el porcentaje de centros educativos que disponen de conexión a Internet; el rendimiento aumenta en un 0,14%. De igual modo, ante un incremento en el porcentaje de alumnos matriculados que cursaron al menos tres años de educación inicial; el rendimiento aumenta en un 0,13%. Por último, en cuanto a la dependencia espacial se concluye que, en la Costa el rendimiento académico en Comprensión Lectora se explica por variables que no están en el modelo de sus regiones vecinas mas si en un 44% en los términos error.

4.2.2.2 Modelación Espacial para la región Sierra.

En este apartado se establecerá qué modelo de regresión espacial para la región Sierra es el más adecuado para el análisis de los resultados.

4.2.2.2.1 Análisis del modelo de la región Sierra.

En primer lugar, se realiza la estimación del modelo para analizar qué determinantes del rendimiento académico en Comprensión Lectora resultan significativos para la región Sierra. Asimismo, se verifica a través del test de Durbin-Wu-Hausman en el que el modelo no presenta problema de endogeneidad (*Anexo N.º 7.16*). Basándose en los resultados de la Tabla N.º 10 se piensa que hay un problema de colinealidad, pues la mayoría de las variables no tienen significancia en el modelo y además algunos signos de los estimadores obtenidos con las pruebas individuales son distintos a los que se espera según la teoría. Sin embargo, a través del factor inflacionario de la varianza se constata que el problema de multicolinealidad es moderada. Por ello, se procede a elegir qué modelo de regresión se debe utilizar para realizar el análisis de los resultados del modelo de la región Sierra.

Tabla N°10. Regresión del modelo de la región Sierra - Pooled Data

A2	Coef.	Z	p > /z/
<i>SB</i>	-0,0008696	-0,83	0,408
<i>ME</i>	0,0002316	0,21	0,830
<i>IN</i>	0,0060652	6,92	0,000
<i>EI</i>	0,001982	4,45	0,000
<i>TC</i>	-0,0077492	-2,39	0,017
<i>PT</i>	-0,0106982	-5,65	0,000
<i>NBI</i>	-0,0003876	0,51	0,613
<i>Cons.</i>	7,201012	31,45	0,000

$N = 45$ $R^2 = 0,83$ $VIF = 2,26$ $Wald\ chi^2 = 177,18$ $Prob > \chi^2 = 0,00$

Elaboración Propia

4.2.2.2.1.1 Elección del modelo óptimo para la región Sierra.

Para elegir el modelo óptimo, es de importancia estimar los Multiplicadores de Lagrange. Como se observa en la Tabla N.º11, los p-value de los modelos de LM-error como el de LM-lag son mayores a 0.05, por lo tanto se acepta la hipótesis nula de ausencia de dependencia espacial, por ello se utilizará el modelo MCO.

Tabla N.º11. Dependencia espacial en regresión MCO

<i>Test</i>	<i>Statics</i>	<i>Df</i>	<i>p – value</i>
<i>Spatial error :</i>			
<i>Lagrange Multiplier</i>	0,223	1	0,637
<i>Robust Lagrange Multiplier</i>	0,056	1	0,814
<i>Spatial lag :</i>			
<i>Lagrange Multiplier</i>	0,187	1	0,665
<i>Robust Lagrange Multiplier</i>	0,020	1	0,889

Elaboración Propia

Una vez determinado que no existe dependencia espacial en la región Sierra, se procederá a elegir qué modelo agrupado, de efectos fijos o de efectos aleatorios se debe utilizar para realizar el análisis de los resultados.

Para la elección del modelo adecuado se utiliza el test de Hausman que permite evaluar la hipótesis nula de que el modelo de efectos aleatorios es el modelo más adecuado, frente a la alternativa de que el modelo sea el de efectos fijos. El resultado de la prueba de Hausman determina un *chi2* de 27.72 y una prob $> chi2$ igual a 0.00 (menor de 0.05), lo que indica que se debe rechazar la hipótesis nula, por lo tanto se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas. A modo de resumen en el *Anexo N.º 7.16* se muestra cómo se eligió el modelo de efectos fijos.

En la Tabla N.º12, se puede observar que tanto la variable de servicios básicos, mobiliario escolar, profesores titulados y el NBI no son significativas para la presente investigación, por lo cual se procede a sacarlos del modelo.

Tabla N.º12. Resumen de Resultados de Coeficientes

<i>Variable</i>		<i>Fe</i>	<i>Re</i>
Main	SB	-0,00350169	-0,00086959
	ME	-0,00079414	0,00023158
	IN	0,00270228*	0,0060652***
	EI	0,00235178*	0,00198202***
	TC	-0,0231684**	-0,00774925*
	PT	-0,00101465	-0,01069818***
	NBI	-0,00121854	0,00038761
	Cons.	6,8149334***	7,2010122***
R-sq		0,26	0,83

$N = 45$ * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Elaboración Propia

4.2.2.2.1.2 Modelo Corregido para la región Sierra.

Como ya se ha seleccionado el modelo de efectos fijos como el óptimo y se ha determinado que se tiene que eliminar del modelo un conjunto de cuatro variables estadísticamente no significativos, el modelo a estimar es el siguiente:

$$\text{Log}(A_{it}) = \beta_0 + \beta_1 * IN_{it} + \beta_2 * EI_{it} + \beta_3 * TC_{it} + \mu_{it}$$

En la Tabla N.º13, se observa que todas las variables incluidas en el último modelo de efectos fijos son significativas al 5% y que *rho* es igual a

0.92, lo que indica que aproximadamente el 92% del error compuesto del modelo se debe a los efectos fijos individuales.

Tabla N°13. Resultados finales de la estimación econométrica

<i>A2</i>	<i>Coef.</i>	<i>T</i>	<i>p > /T/</i>
<i>IN</i>	0,0014087	1,44	0,043
<i>EI</i>	0,0025586	3,58	0,001
<i>TC</i>	-0,0188549	-2,67	0,012
<i>Cons.</i>	6,444767	49,73	0,000

$$N = 45 \quad R^2 = 0,26 \quad \text{Wald } \chi^2 = 4,65 \quad \text{Prob} > \chi^2 = 0,86 \quad \rho = 0,92$$

Elaboración Propia

Quedando como modelo final:

$$\text{Log}(A_{it}) = 6.44 + 0.0014 * IN_{it} + 0.0026 * EI_{it} - 0.019 * TC_{it} + \mu_i$$

En la región Sierra, los determinantes del rendimiento académico en Comprensión Lectora son el acceso a internet, la educación inicial y el tamaño de clase. Ante un incremento en el porcentaje de centros educativos que disponen de conexión a Internet; el rendimiento aumenta en un 0,14%. Asimismo, ante un incremento en el porcentaje de alumnos matriculados que cursaron al menos tres años de educación inicial; el rendimiento aumenta en un 0,26%. De igual modo, ante el incremento de un alumno por clase; el rendimiento disminuye en un 1,9%.

4.2.2.3 Modelación Espacial para la región Selva.

En este apartado se establecerá qué modelo de regresión espacial para la región Selva es el más adecuado para el análisis de los resultados.

4.2.2.3.1 Análisis del modelo para la región Selva.

En primer lugar, se realiza la estimación del modelo para analizar qué determinantes del rendimiento académico en Comprensión Lectora resultan significativos para la región Selva. Asimismo, se verifica a través del test de Durbin-Wu-Hausman en el que el modelo no presenta problema de endogeneidad (*Anexo N.º 7.17*). Basándose en los resultados de la Tabla N.º 14, se piensa que hay un problema de colinealidad, pues la mayoría de las variables no tienen significancia en el modelo y además algunos signos de los estimadores obtenidos con las pruebas individuales son distintos a los que se espera según la teoría. Sin embargo, a través del factor inflacionario de la varianza se constata que el problema de multicolinealidad es moderada. Por ello, se procede a elegir qué modelo de regresión se debe utilizar para realizar el análisis de los resultados del modelo de la región Selva.

Tabla N°14. Regresión del modelo para la región Selva - Pooled Data

A2	Coef.	Z	p > z/
<i>SB</i>	-0,0018693	2,64	0,008
<i>ME</i>	-0,0002908	-0,21	0,833
<i>IN</i>	0,0022875	2,00	0,045
<i>EI</i>	0,0027168	4,26	0,000
<i>TC</i>	-0,0119556	-1,23	0,217
<i>PT</i>	0,0067676	3,33	0,001
<i>NBI</i>	3,22 e-06	0,00	0,999
<i>Cons.</i>	5,74924	36,01	0,000

$N = 40$ $R^2 = 0,91$ $Wald\ chi^2 = 172,64$ $Prob > \chi^2 = 0,00$

Elaboración Propia

4.2.2.3.1.1 Elección del modelo óptimo para la región Selva.

Para elegir el modelo óptimo, es de importancia estimar los Multiplicadores de Lagrange. Como se observa en la Tabla N.º15, los p-value de los modelos de LM-error como el de LM-lag son mayores a 0.05, por lo tanto se acepta la hipótesis nula de ausencia de dependencia espacial, por ello se utilizará el modelo MCO.

Tabla N.º15. Dependencia espacial en regresión MCO

<i>Test</i>	<i>Statics</i>	<i>Df</i>	<i>p – value</i>
<i>Spatial error :</i>			
<i>Lagrange Multiplier</i>	1,629	1	0,202
<i>Robust Lagrange Multiplier</i>	5,417	1	0,020
<i>Spatial lag :</i>			
<i>Lagrange Multiplier</i>	0,022	1	0,882
<i>Robust Lagrange Multiplier</i>	3,810	1	0,051

Elaboración Propia

Una vez determinado que no existe dependencia espacial en la región Selva, se procederá a elegir qué modelo agrupado, de efectos fijos o de efectos aleatorios se debe utilizar para realizar el análisis de los resultados.

Para la elección del modelo adecuado se utiliza el test de Hausman que permite evaluar la hipótesis nula de que el modelo de efectos aleatorios es el modelo más adecuado, frente a la alternativa de que el modelo sea el de efectos fijos. El resultado de este test determina un χ^2 de 51.39 y una prob $> \chi^2$ igual a 0.00 (menor de 0.05), lo que indica que se debe rechazar la hipótesis nula, por lo tanto se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas. A modo de resumen en el Anexo N.º 7.17 se muestra cómo se eligió el modelo de efectos fijos.

En la Tabla N.º16, se puede observar que tanto la variable de servicios básicos, mobiliario escolar, acceso a Internet, tamaño de clase y profesores titulados no son significativas para la presente investigación, por lo cual se procede a sacarlos del modelo.

Tabla N.º16. Resumen de Resultados de Coeficientes

<i>Variable</i>		<i>Fe</i>	<i>Re</i>
Main	SB	0,00088413	0,00186928**
	ME	-0,00123641	-0,00029077
	IN	0,00143645	0,00228746*
	EI	0,00334071***	0,00271681***
	TC	0,00582771	-0,01195564
	PT	-0,00164983	0,00676758***
	NBI	0,00568999*	3,218 e-06
	Cons.	5,9578854***	5,7492398***
R-sq		0,05	0,91

$N = 40$ * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

4.2.2.3.1.2 Modelo Corregido para la región Selva

Como ya se ha seleccionado el modelo de efectos fijos como el óptimo y se ha determinado que se tiene que eliminar del modelo un conjunto de cinco variables estadísticamente no significativos, el modelo a estimar es el siguiente:

$$\text{Log}(A_{it}) = \beta_0 + \beta_1 * EI_{it} + \beta_2 * NBI_{it} + \mu_{it}$$

En la Tabla N.º17, se observa que todas las variables incluidas en el último modelo de efectos fijos son significativas al 5% y que ρ es igual a 0.98, lo que indica que aproximadamente el 98% del error compuesto del modelo se debe a los efectos fijos individuales.

Tabla N.º17. Resultados finales de la estimación econométrica

<i>A2</i>	<i>Coef.</i>	<i>T</i>	<i>p > T </i>
<i>EI</i>	0,0039214	10,95	0,000
<i>NBI</i>	-0,0131852	1,91	0,049
<i>Cons.</i>	5,886416	51,70	0,000

$$N = 40 \quad R^2 = 0,4 \quad \text{Wald } \chi^2 = 2,29 \quad \text{Prob} > \chi^2 = 0,81 \quad \rho = 0,98$$

Elaboración Propia

Quedando como modelo final para la región Selva:

$$\text{Log}(A_{it}) = 5.89 + 0.0039 * EI_{it} - 0.013 * NBI_{it} + \mu_i$$

En la región Selva, los determinantes del rendimiento académico en Comprensión Lectora son la educación inicial y las Necesidades Básicas Insatisfechas. Ante un incremento en el porcentaje de alumnos matriculados que cursaron al menos tres años de educación inicial; el rendimiento aumenta en un 0,39%. De igual modo, ante el incremento del porcentaje de personas que cuenten con al menos una Necesidad Básica Insatisfecha; el rendimiento disminuye en un 1,3%.

4.2.3 Análisis de los Resultados según ámbito geográfico.

Los resultados obtenidos de los modelos específicos sugieren que los determinantes del rendimiento académico en Comprensión Lectora según el ámbito geográfico son diferentes, aceptando de esta manera la hipótesis específica del trabajo de investigación. Estas diferencias en los determinantes del rendimiento se deben principalmente a la inequidad existente entre las regiones (Consejo Nacional de Educación, 2015).

Asimismo, se resalta que los modelos de regresión que no incluyen las variables estadísticamente no significativas, son útiles en caso se quiera aplicar modelos predictivos del rendimiento académico, según ámbito geográfico. Aunque este no sea un objetivo del trabajo de investigación, es importante mencionarlo.

Los resultados sugieren que los determinantes del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 son los siguientes:

- Costa: Acceso a Internet y Educación Inicial.
- Sierra: Acceso a Internet, Educación Inicial y Tamaño de clase.
- Selva: Educación Inicial y las Necesidades Básicas Insatisfechas.

Por otro lado, los rendimientos académicos de la Costa son mayores en comparación con las regiones de la Sierra y Selva. Estos resultados podrían explicarse por la oferta educativa en cada región geográfica. Asimismo, se debe tener en cuenta que “gestionar los recursos para la región Selva es un desafío debido al territorio extenso, disperso y desarticulado sobre todo si no existe una institucionalidad política regional sólida” (Christiansen, Garret, & Marcos, 2016, p. 4). De esta manera, se confirma lo observado en el cuadro del *Anexo N.º 7.11.8*, dónde la Selva tiene mayor población con al menos una Necesidad Básica Insatisfecha, seguido de la Sierra y Costa lo cual afecta inversamente sobre el rendimiento académico.

CAPITULO V: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES

La presente investigación nos permite cumplir con los objetivos de la investigación, los cuales son identificar los determinantes del rendimiento académico en las regiones del Perú y determinar la existencia de dependencia espacial del rendimiento académico. Para ello se utilizó la base de datos proporcionada por la Estadística de Calidad Educativa (ESCALE 2012-2016) y las Evaluaciones Censales de Estudiantes (ECE), las cuales proporcionaron la información sobre los indicadores de calidad educativa y los resultados del rendimiento académico de las regiones del Perú durante los años estudiados. Asimismo, es importante señalar que se toma la variable de rendimiento académico en Comprensión Lectora como variable proxy del rendimiento académico, lo cual genera una limitación en el trabajo de investigación.

En cuanto a la dependencia espacial se concluye a nivel general que, el rendimiento académico de una de una región influye en un 80% al rendimiento educativo de sus colindantes agrupándose por ubicación geográfica. Es decir, los resultados del rendimiento académico se agrupan según las regiones geográficas donde se observa que las regiones de la Costa tienen un rendimiento académico más alto que las regiones de la Sierra y Selva.

A nivel general, las variables de servicios básicos, educación inicial y tamaño de clase tienen estadísticos significativos para medir el rendimiento académico en comprensión lectora. Por otro lado, con respecto a las regiones geográficas, los determinantes del rendimiento académico que resultaron tener estadísticos significativos fueron: para la Costa, las variables de acceso a Internet y educación inicial; para la Sierra, las variables de acceso a Internet, educación inicial y el tamaño de clase; finalmente para la Selva las variables que resultaron tener estadísticos significativos fueron la educación inicial y las Necesidades Básicas Insatisfechas.

En conclusión, a nivel general los servicios básicos son esenciales para el buen rendimiento académico en comprensión lectora. Pues el contar con energía eléctrica no solo permite que los alumnos estudien más, sino que también tengan acceso a Internet, que resulta ser una variable significativa para el rendimiento académico de la Costa y Sierra (MINEDU, 2017a). De igual forma se concluye que a menor número de estudiantes por clase hay un mayor rendimiento académico, pues la educación es más personalizada (Ministerio de Educación, Cultura y Deporte, 2016). Con respecto a la variable de Necesidad Básica Insatisfecha se concluye que las regiones menos pobres como la Costa muestran mejores

rendimientos que las regiones de la Sierra y Selva (Ramírez, Quiñones, Erasmo, Salazar, & Aníbal, 2011).

Analizando el rendimiento académico según ámbito geográfico, se sugiere que existe jerarquía en los determinantes del rendimiento académico, según su importancia. Por ello, el Gobierno debería implementar políticas públicas que logren reducir la brecha de desigualdad del rendimiento académico en las regiones. En primer lugar, a nivel general se resalta la importancia que tiene la educación inicial pues no sólo es útil para el rendimiento académico futuro de los estudiantes sino también sirve como medio para eliminar las desigualdades regionales, por ello el Ministerio de Educación debería implementar proyectos para aumentar la cobertura y calidad de la educación inicial para todas las regiones. En segundo lugar, para la región Selva se resalta la importancia de las Necesidades Básicas Insatisfechas, por ello el Ministerio de Vivienda, Construcción y Saneamiento debería sumar esfuerzos para llegar no sólo a las zonas urbanas sino también a las zonas rurales de las regiones de la Selva, con el objetivo de que todas familias puedan vivir en un hogar sin Necesidades Básicas Insatisfechas pues este determinante resulta fundamental para el rendimiento académico de los alumnos de la Selva. Por último, para las regiones Sierra y Costa se resalta la importancia del tamaño promedio de clase y el acceso al Internet respectivamente, por ello el Ministerio de Educación no solo debería de hacer cumplir sus leyes en cuánto al mínimo número de estudiantes en un aula sino también implementar políticas para la cobertura del acceso a internet en todos los colegios ya sea urbano rural.

Los hallazgos de este estudio indican que la manera en que el sistema educativo se configura en las regiones de la Costa, Sierra y Selva son distintas, lo que genera diferencias en los resultados académicos de los estudiantes por lo cual se recomienda realizar más estudios de investigación haciendo uso de la econometría espacial, ya que las ubicaciones de las regiones de por sí, tienen diferencias tanto culturales como geográficas que pueden influir en el rendimiento académico. Asimismo, en futuras investigaciones se recomienda incluir más determinantes que permitan medir el rendimiento académico desde el lado de la demanda como por ejemplo la lengua materna y el nivel educativo de los padres, ya que los determinantes empleados por esta investigación están enfocados más desde el lado de la oferta. De igual modo, sería bueno investigar las diferencias que existen en el rendimiento académico entre llevar clases virtuales y clases presenciales, comparando los resultados académicos del 2019 contra el 2020, pues todos los estudiantes de primaria llevaron clases virtuales en este año.

BIBLIOGRAFÍA

- Acevedo, I., & Velásquez, E. (Octubre de 2008). *Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales*. Obtenido de [237769553_Algunos_conceptos_de_la_econometria_espacial_y_el_analisis_exploratorio_de_datos_espaciales](https://doi.org/10.237769553_Algunos_conceptos_de_la_econometria_espacial_y_el_analisis_exploratorio_de_datos_espaciales)
- Anselin, L. (1995). *New Directions in Spatial Econometrics*. Obtenido de [https://books.google.com.pe/books?hl=es&lr=&id=Oef7CAAAQBAJ&oi=fnd&pg=PR14&dq=Anselin,+L.+\(+1995+\).pdf&ots=LLsloGGQng&sig=-sBuJ21MbSCiG-dBpZ6ic8qHTzU#v=onepage&q&f=false](https://books.google.com.pe/books?hl=es&lr=&id=Oef7CAAAQBAJ&oi=fnd&pg=PR14&dq=Anselin,+L.+(+1995+).pdf&ots=LLsloGGQng&sig=-sBuJ21MbSCiG-dBpZ6ic8qHTzU#v=onepage&q&f=false)
- Anselin, L. (2005). *Explorind Spatial Data with GeoDa*. Obtenido de <http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf>
- BANCO MUNDIAL. (2018). *Gasto público en educación, total (%PBI)*. Obtenido de <https://datos.bancomundial.org/indicador/SE.XPD.TOTL.GD.ZS>
- Barnett, W., & Esposito, C. (2006). *Impactos estimados de la cantidad de años de asistencia preescolar en vocabulario, alfabetización y habilidades matemáticas en la entrada al jardín de infantes*. Obtenido de https://www.researchgate.net/publication/253316056_Estimated_Impacts_of_Number_of_Years_of_Preschool_Attendance_on_Vocabulary_Literacy_and_Math_Skills_at_Kindergarten_Entry
- Becker, G. (1964). *Human Capital*. Obtenido de <https://www.econstor.eu/bitstream/10419/152338/1/815411251.pdf>
- Beltrán, A., & Seinfeld, J. (2012). *La trampa educativa en el Perú*. Lima: Universidad del Pacífico.
- Berlinski, S., Galiani, S., & Gertler, P. (2006). *El efecto de la educación pre-primaria en el rendimiento escolar primaria*. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0047272708001308?via%3Dihub>
- Borrego, J. (2018). *Modelos de Regresión para Datos Espaciales*. Obtenido de <https://idus.us.es/bitstream/handle/11441/81660/Borrego%20S%E1nchez%20Jo>

s%E9%20%C1ngel%20TFG.pdf;jsessionid=2057F720C8FBE704030264307C1AEBD5?sequence=1&isAllowed=y

- Caballero, C., Abello, R., & Palacio, J. (2007). *Relación de burnout y rendimiento académico con la satisfacción frente a los estudios en estudiantes universitarios. Avances en Psicología Latinoamericana*. Bogota: Redalyc.
- Campana, Y., Velasco, D., Aguirre, J., & Guerrero, E. (2014). *Inversión en infraestructura educativa: una aproximación a la medición de sus impactos a partir de la experiencia de los Colegios Emblemáticos*. Lima: Consorcio de investigación económica y social.
- Cardona, M., Montes, I., Vásquez, J., Villegas, M., & Brito, T. (Abril de 2007). *Capital Humano: Una mirada desde la educación y la experiencia laboral*. Obtenido de <http://publicaciones.eafit.edu.co/index.php/cuadernos-investigacion/article/view/1287/1166>
- Carrasco, S. (2005). *Metodología de la Investigación Científica : pautas metodológicas para diseñar y elaborar el proyecto de investigación*. Lima: San Marcos.
- Celemín, J. (2009). *Autocorrelación espacial e indicadores locales de asociación espacial*. Obtenido de <https://www.redalyc.org/pdf/3832/383239099001.pdf>
- Cliff, A., & Ord, J. (1973). *Classics in human geography revisited*. Obtenido de <https://journals.sagepub.com/doi/abs/10.1177/030913259501900205?journalCode=phgb>
- Coleman, J. (1966). *La posibilidad de una función de bienestar social*. Obtenido de <https://www.jstor.org/stable/1815299?seq=1>
- Consejo Nacional de Educación. (2015). *Proyecto educativo nacional: balance y recomendaciones 2015*. Obtenido de file:///D:/BALANCE-PEN-2015_02_final.pdf
- Correa, S., Correa, A., & Alvarez, A. (s.f.). *La gestión educativa un nuevo paradigma*. Medellín.
- Cueto, S. (2007). *Las Evaluaciones Nacionales e Internacionales de Rendimiento Escolar en el Perú : Balances y Perspectivas*. Lima. Obtenido de

<http://biblioteca.clacso.edu.ar/Peru/grade/20100513025532/InvPolitDesarr-10.pdf>

Chasco, C. (Enero de 2004). *Modelos de heterogeneidad espacial*. Obtenido de https://www.researchgate.net/publication/23742953_Modelos_de_heterogeneidad_espacial

Christiansen, A., Garret, P., & Marcos, M. (2016). *Regiones en perspectiva: la influencia de los factores asociados al aprendizaje al término de la educación primaria*. Obtenido de http://umc.minedu.gob.pe/wp-content/uploads/2016/07/EB02_Regiones_en_perspectiva_VF.pdf

Currie, J., & Thomas, D. (2000). *School Quality and the longer-term effect of head start*. Obtenido de <https://www.nber.org/papers/w6362.pdf>

Dall'erba, S., & Le Gallo, J. (2004). *Regional convergence and the impact of european structural fund over 1989-1999*. Obtenido de https://www.researchgate.net/publication/4780678_Regional_convergence_and_the_impact_of_European_structural_funds_over_1989-1999_A_spatial_econometric_analysis

Deninson, E. (1962). *The sources of Economic Growth in the United States*. Obtenido de <https://academic.oup.com/ej/article-abstract/72/288/935/5249393?redirectedFrom=fulltext>

Elias, M., & Rey, S. (2011). *Educational Performance and Spatial Convergence in Peru*. Obtenido de <http://region-developpement.univ-tln.fr/en/pdf/R33/Elias.pdf>

Elías, S. (Enero de 2004). *Capital humano, calidad educativa y crecimiento económico*. Obtenido de http://bibliotecadigital.uns.edu.ar/scielo.php?script=sci_arttext&pid=S2525-12952004001100002&lng=en&nrm=iso

ESCALE. (2018). Obtenido el 2020, de <http://escale.minedu.gob.pe/indicadores?jsessionid=b561a488efa385eab7b6a06f1665>

- Escobar, F. (2006). *Importancia de la educación inicial a partir de la mediación de los procesos cognitivos para el desarrollo humano integral*. Caracas: Laurus.
- Filmer, D. (2003). *Determinants of Health and Education Outcomes*.
- Fuentes, C., Tarazona, L., Tarquino, F., & Villanueva, C. (2016). *Análisis, evaluación y propuesta de un modelo de indicadores de eficiencia para la educación primaria*. Lima.
- Fullan, M., Schleicher, A., Lee, S., Gopinathan, S., & Hill, P. (2007). *How the world's best-performing school systems come out on top*. Obtenido de https://www.mckinsey.com/~media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/How%20the%20worlds%20best%20performing%20school%20systems%20come%20out%20on%20top/How_the_world_s_best-performing_school_systems_come_out_on_top.pdf
- Getis, A. (2010). *Spatial Autocorrelation*. Obtenido de https://link.springer.com/chapter/10.1007/978-3-642-03647-7_14
- Glewwe, P., & Kremer, M. (2006). Schools, Teachers, and Education outcomes in developing countries. En *Handbook of the Economics of Education* (págs. 945-1017). Elsevier B.V.
- Hanushek, E. (1989). *The Impact of Differential Expenditures on School Performance*. American Educational Research Association. Obtenido de <http://hanushek.stanford.edu/sites/default/files/publications/hanushek%201989%20EducResearcher%2018%284%29.pdf>
- Hanushek, E., & Wößmann, L. (2007). *The Role of Education Quality in Economic Growth*. Obtenido de <http://documents1.worldbank.org/curated/en/260461468324885735/pdf/wps4122.pdf>
- Hendry, D. (2000). *Econometría: ¿alquimia o ciencia?* Oxford. Obtenido de <https://www.oxfordscholarship.com/view/10.1093/0198293542.001.0001/acprof-9780198293545-chapter-12>

- Hernández, B. (2011). *El mobiliario y su funcionalidad en la Escuela Oficial Urbana Mixta del Municipio de Atescatempa, Departamento de Jutiapa*.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación*. México D.F.: MC GRAW - HILL / INTERAMERICANA EDITORES , S.A. DE C.V.
- Hincapié, G. (2017). *Efectos Contagio sobre el Rendimiento Educativo Municipal del Gasto Público en Educación en el departamento de Antioquia*. Medellín. Obtenido de <http://bdigital.unal.edu.co/55770/>
- Hsiao, C. (2003). *Analysis of panel data* . Cambridge: Cambridge University.
- INEE. (2014). *Mobiliario y equipo básico para la enseñanza y el aprendizaje* . México D.F.
- INEI. (2018). Obtenido el 2020, de <https://www.inei.gob.pe/estadisticas/indice-tematico/sociales/>
- Internet Society . (2017). *Acceso a Internet y educación:Consideraciones clave para legisladores* .
- IPE. (2018). *Indice de Competitividad Regional - 2018*. Lima. Obtenido de <https://www.ipe.org.pe/portal/incore-2018-indice-de-competitividad-regional/>
- Jadue, G. (1997). *Factores ambientales que afectan el rendimiento escolar de los niños provenientes de familias de bajo nivel socioeconómico y cultural* . Obtenido de https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-07051997000100007
- Lesage, J., & Pace, R. (2009). *Introduction to Spatial Econometrics*. Florida: Chapm& Hall/ CRC Press. Obtenido de <https://core.ac.uk/download/pdf/81632768.pdf>
- Lucas, R. (1988). *On the Mechanics of Economic Development*. Obtenido de <http://kisi.deu.edu.tr/yesim.kustepeli/uzawa1965.pdf>
- Mankiw, N., Romer, D., & Weil, D. (1992). *A Contribution to the Empirics of Economic Growth*. Obtenido de <http://piketty.pse.ens.fr/files/MankiwEtal92.pdf>

- Martínez-Otero, V. (2007). *Los adolescentes ante el estudio. Causas y consecuencias del rendimiento académico*. Madrid: Fundamentos.
- MEF. (s.f.). *Métodos para medir Pobreza*. Obtenido el 2020, de <https://www.mef.gob.pe/es/mapas-de-pobreza/metodos-para-medir-la-pobreza>
- MINEDU. (2016a). *Ley General de la Educación*. Lima. Obtenido de <http://www.minedu.gob.pe/comunicado/pdf/normativa-2018/ley-28044/ds-011-2012-24-11-2017.pdf>
- MINEDU. (2016b). *Regiones en perspectiva : la influencia de los factores asociados al aprendizaje al término de la educación primaria*. Obtenido de http://umc.minedu.gob.pe/wp-content/uploads/2016/07/EB02_Regiones_en_perspectiva_VF.pdf
- MINEDU. (2017a). *¿Cómo se relaciona la infraestructura de la escuela con los aprendizajes de los estudiantes?* Lima: Oficina de Medición de la Calidad de los Aprendizajes.
- MINEDU. (2018). *Evaluación PISA 2018*. Obtenido de http://umc.minedu.gob.pe/wp-content/uploads/2020/09/PPT-PISA-2018_Web_vf-31-08-20.pdf
- Ministerio de Educación, Cultura y Deporte. (2016). *Panorama de la educación. Indicadores de la OCDE 2016*. Madrid.
- Montero, R. (Junio de 2011). *Efectos fijos o aleatorios: test de especificación*. Obtenido de <https://www.ugr.es/~montero/matematicas/especificacion.pdf>
- Moreno, R., & Vayá, E. (2004). *Técnicas econométricas para datos espaciales*. Obtenido de https://books.google.com.pe/books?id=Udh_wcm75GwC&pg=PA14&lpg=PA14&dq=T%C3%A9cnicas+econ%C3%B3micas+para+el+tratamiento+de+datos+espaciales:+la+econometr%C3%ADa+espacial&source=bl&ots=9G4NR8oIVf&sig=ACfU3U0cD-jQ3fvwa3oZGqf_H_-sjcL5Nw&hl=es-419&sa=X&ved=2a
- Mosteller, F. (1995). *The Tennessee study of class size in the early school grades*.

- Myers, R. (1992). *Los doce que sobreviven: Fortalecimiento de los programas de desarrollo de la primera infancia en el Tercer Mundo*. Obtenido de <https://www.journals.uchicago.edu/doi/pdfplus/10.1086/447279>
- Naidoo, Eeden, & Munch. (2014). *Spatial Variation in School Performance, a Local Analysis of Socio-economic Factors in Cape Town*. Obtenido de <https://www.ajol.info/index.php/sajg/article/view/106138>
- Niu, S., Jia, Y., Wang, W., He, R., Hu, L., & Liu, Y. (2013). *Electricity consumption and human development level: A comparative analysis based on panel data for 50 countries*. Lanzhou: Elsevier.
- Otero, M. (2006). *La escuela impacta más en los pobres*. Obtenido de http://archivo.lavoz.com.ar/06/10/08/secciones/sociedad/nota.asp?nota_id=7274
- Paelinck, J., Mur, J., & Trívez, F. (2004). *Econometría Espacial : más luces que sombras*. Obtenido de https://www.researchgate.net/publication/5164596_Econometria_espacial_mas_luces_que_sombras
- Palacios, N. (2009). *Los efectos de la educación en el crecimiento de América Latina*. Bogotá. Obtenido de <https://repository.urosario.edu.co/bitstream/handle/10336/1469/PalaciosVera-Natalia-2009.pdf;jsessionid=8EC40440DCE5FCD3045FB9A9C09FCE67?sequence=1>
- Pérez, J. (2006). *Econometría espacial y ciencia regional*. Obtenido de <https://www.redalyc.org/pdf/601/60125804.pdf>
- Rajimon, J. (2010). *La economía y la función de producción en educación*. Obtenido de <https://www.redalyc.org/pdf/3579/357935475006.pdf>
- Ramírez, M., Quiñones, D., Erasmo, R., Salazar, L., & Aníbal, R. (2011). *Pobreza y rendimiento escolar: Estudio de caso de jóvenes de alto rendimiento*. Obtenido de <https://www.redalyc.org/pdf/356/35622379013.pdf>

- Rodríguez, R. (2015). *La educación: elemento clave en la reducción de la pobreza y la desigualdad*. Obtenido de http://www.unescoetxea.org/dokumentuak/CursoODS2015_18_educacionreduccionpobreza.pdf
- Sánchez, M. (2012). *El papel de la familia en la educación*. Albacete.
- Schultz, T. (Marzo de 1961). *Investment in Human Capital*. Obtenido de <http://la.utexas.edu/users/hcleaver/330T/350kPEESchultzInvestmentHumanCapital.pdf>
- SICRECE. (2018). Obtenido el 2020, de https://sistemas15.minedu.gob.pe:8888/evaluacion_censal_publico
- Tobler, W. (1970). *A Computer Movie Simulating Urban Growth in the Detroit Region*. Obtenido de <http://www.scielo.org.co/pdf/rcdg/v28n1/2256-5442-rcdg-28-01-1.pdf>
- Tomasevski, K. (2004). *Indicadores al derecho de educación*. Obtenido de <http://www.derechoshumanos.unlp.edu.ar/assets/files/documentos/indicadores-del-derecho-a-la-educacion.pdf>
- UNESCO. (2007). *La formación de un docente de calidad para el desarrollo sostenible*. Obtenido de http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Havana/pdf/Formaciondocentes_Llivina.pdf
- UNICEF. (2017). *La infancia paga el precio de la falta de agua potable y saneamiento*. Madrid. Obtenido de <https://www.unicef.es/prensa/la-infancia-paga-el-precio-de-la-falta-de-agua-potable-y-saneamiento>
- Uzawa, H. (1965). *Optimum Technical Change in An Aggregative Model of Economic Growth*. Obtenido de <http://kisi.deu.edu.tr/yesim.kustepeli/uzawa1965.pdf>
- Zuniga, P., Ordoñez, A., Ordoñez, V., & Martínez, C. (2018). *Educación y crecimiento económico: Análisis e Implicancias*. Obtenido de <https://www.lamjol.info/index.php/EyA/article/view/6654>

ANEXOS

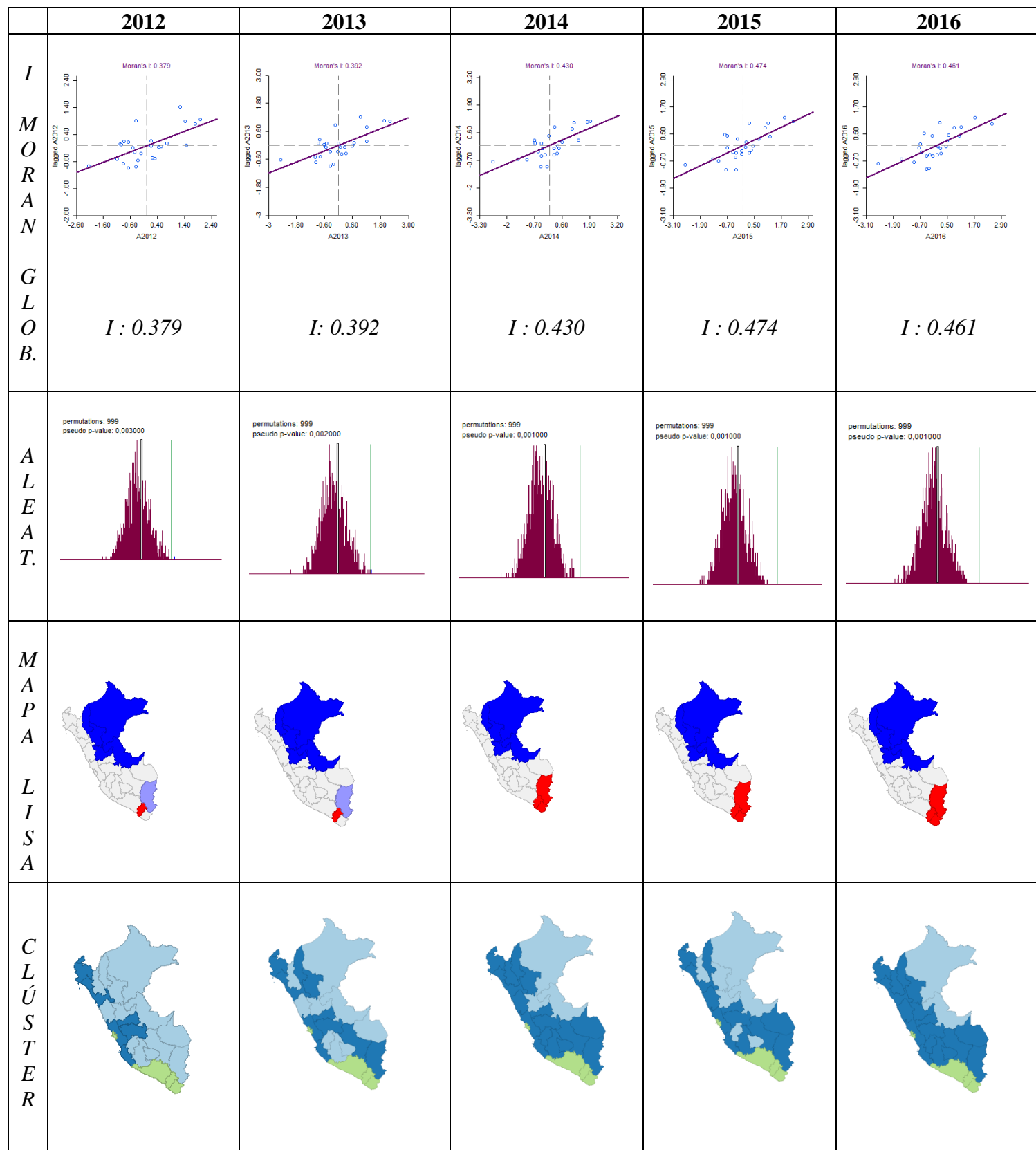
7.1 Gasto Público en Educación

Tabla N° 28. *Gasto público en Educación del 2010 - 2017 (% PBI)*

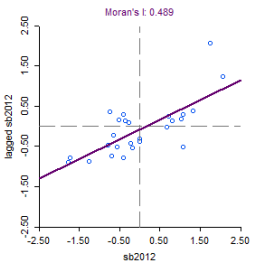
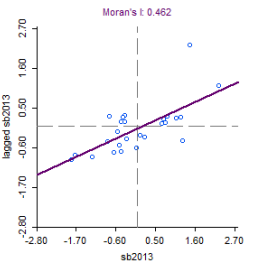
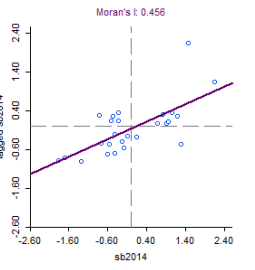
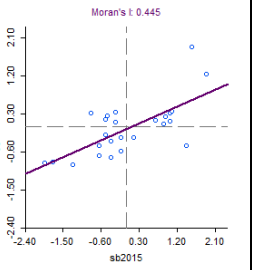
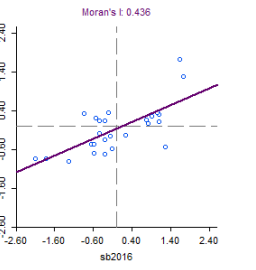
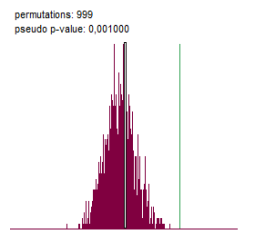
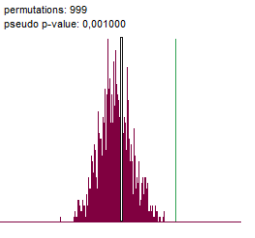
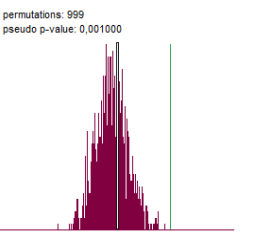
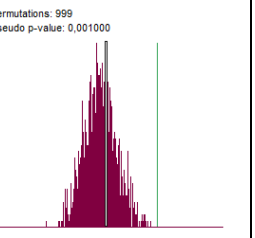
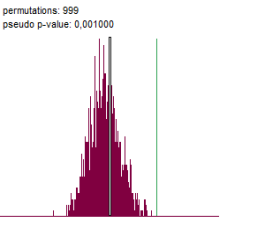






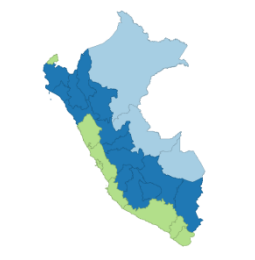

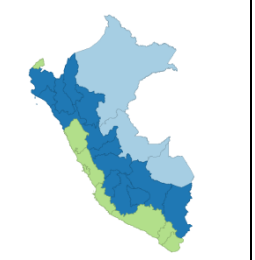
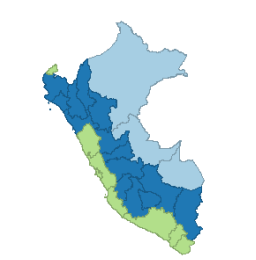
<i>Países</i>	<i>2010</i>	<i>2011</i>	<i>2012</i>	<i>2013</i>	<i>2014</i>	<i>2015</i>	<i>2016</i>	<i>2017</i>
<i>Nueva Zelandia</i>	7	7	7,2	6,7	6,4	6,3	6,3	6,3
<i>Brasil</i>	5,7	5,7	5,9	6	6	6,2	6,2	6,2
<i>Estados Unidos</i>	5,4	5,2	5,2	4,9	5	5	5	5
<i>Argentina</i>	5	5,3	5,4	5,4	5,4	5,8	5,6	5,5
<i>Colombia</i>	4,8	4,5	4,4	4,9	4,7	4,5	4,5	4,4
<i>Tailandia</i>	3,5	4,8	4,6	4,1	4,1	4,3	4,1	4,1
<i>Ecuador</i>	4,5	4,7	4,6	5	5,3	5	5	5
<i>Chile</i>	4,2	4,1	4,6	4,6	4,8	4,9	5,4	5,4
<i>El Salvador</i>	3,5	3,4	3,3	3,4	3,4	3,5	3,7	3,9
<i>Perú</i>	2,9	2,7	2,9	3,3	3,7	4	3,8	3,9

Fuente: Banco Mundial - Elaboración: Propia

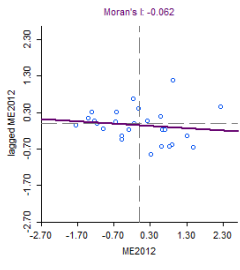
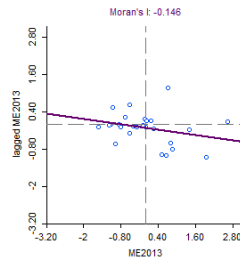
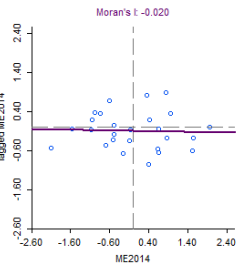
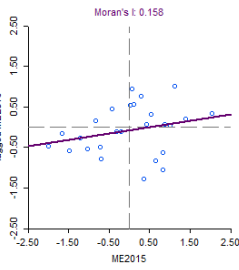
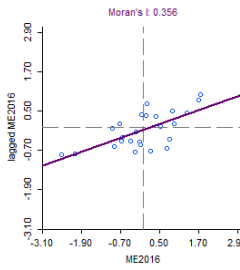
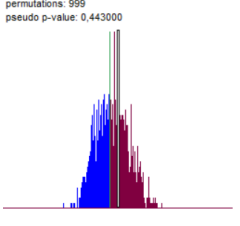
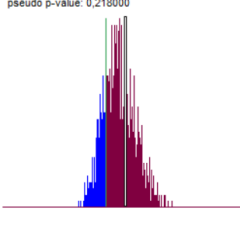
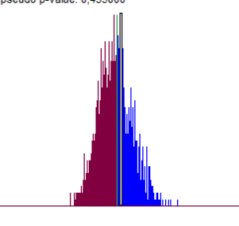
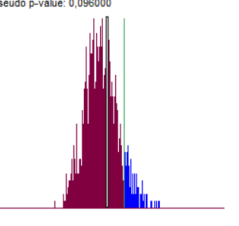
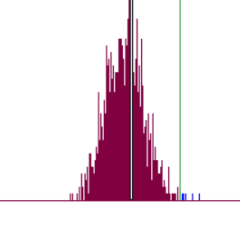




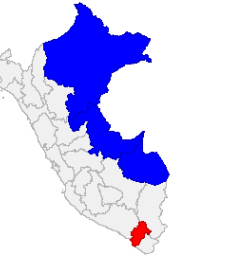
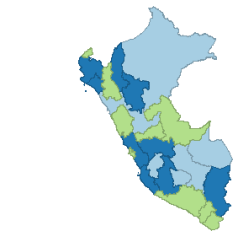
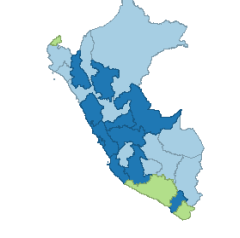
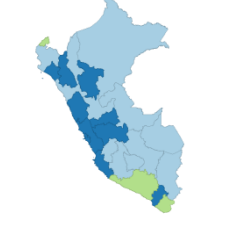
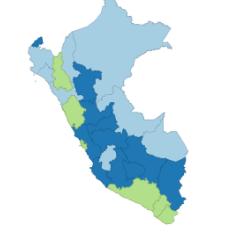
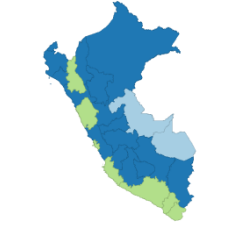
7.2 AEDE del Rendimiento Académico.



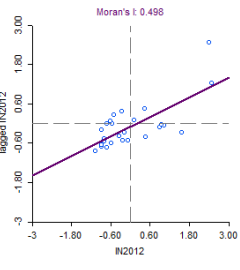
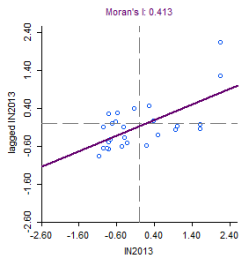
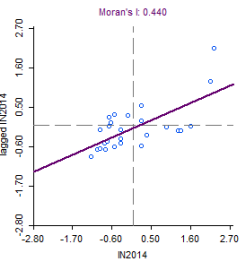
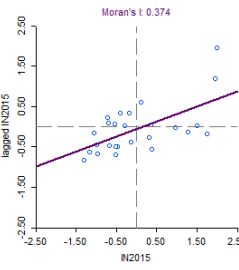
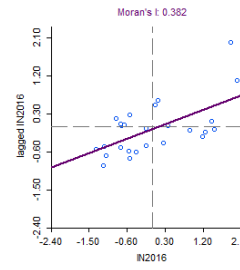
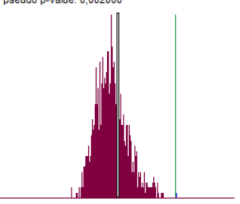
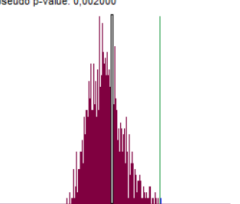
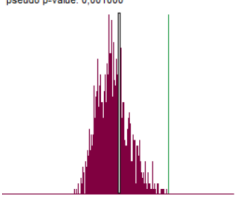
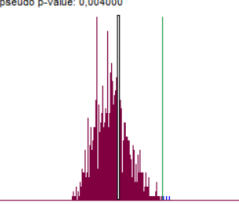
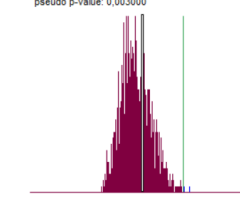









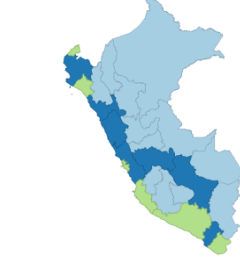
7.3 AEDE del Servicios Básicos.

	2012	2013	2014	2015	2016
I M O R A N G L O B.	 <p>Moran's I: 0.489</p>	 <p>Moran's I: 0.462</p>	 <p>Moran's I: 0.456</p>	 <p>Moran's I: 0.445</p>	 <p>Moran's I: 0.436</p>
A L E A T.	 <p>permutations: 999 pseudo p-value: 0,001000</p>	 <p>permutations: 999 pseudo p-value: 0,001000</p>	 <p>permutations: 999 pseudo p-value: 0,001000</p>	 <p>permutations: 999 pseudo p-value: 0,001000</p>	 <p>permutations: 999 pseudo p-value: 0,001000</p>
M A P A L I S A					
C L Ú S T E R					

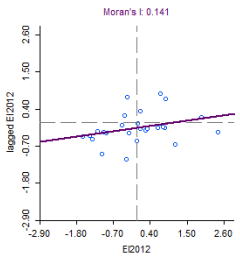
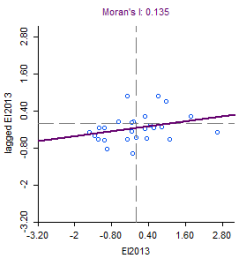
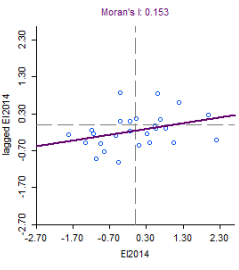
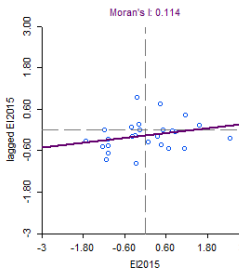
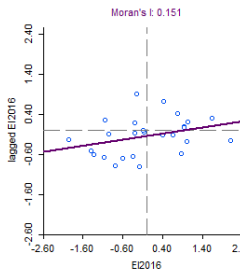
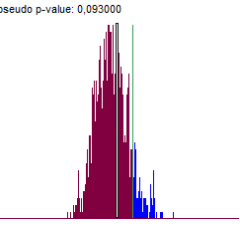
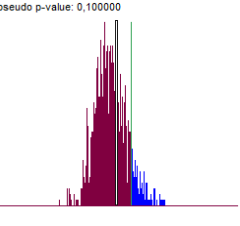
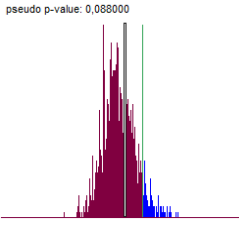
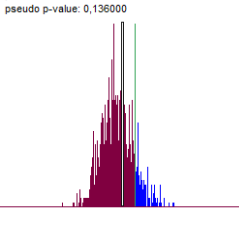
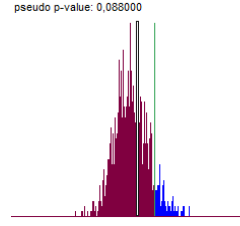








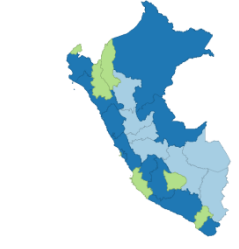
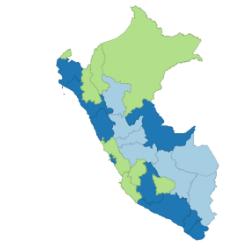
7.4 AEDE del Mobiliario Escolar.

	2012	2013	2014	2015	2016
<i>I</i> <i>M</i> <i>O</i> <i>R</i> <i>A</i> <i>N</i> <i>G</i> <i>L</i> <i>O</i> <i>B</i>	 <p>Moran's I: -0.062</p> <p>$I: -0.062$</p>	 <p>Moran's I: -0.146</p> <p>$I: -0.146$</p>	 <p>Moran's I: -0.020</p> <p>$I: -0.02$</p>	 <p>Moran's I: 0.158</p> <p>$I: 0.158$</p>	 <p>Moran's I: 0.356</p> <p>$I: 0.356$</p>
<i>A</i> <i>L</i> <i>E</i> <i>A</i> <i>T</i> .	 <p>permutations: 999 pseudo p-value: 0.443000</p>	 <p>permutations: 999 pseudo p-value: 0.218000</p>	 <p>permutations: 999 pseudo p-value: 0.435000</p>	 <p>permutations: 999 pseudo p-value: 0.096000</p>	 <p>permutations: 999 pseudo p-value: 0.007000</p>
<i>M</i> <i>A</i> <i>P</i> <i>A</i> <i>L</i> <i>I</i> <i>S</i> <i>A</i>					
<i>C</i> <i>L</i> <i>Ú</i> <i>S</i> <i>T</i> <i>E</i> <i>R</i>					

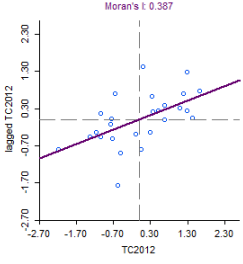
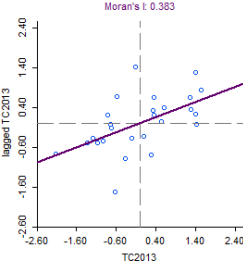
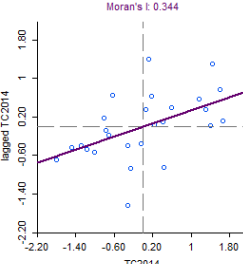
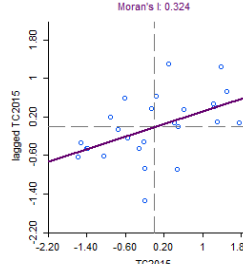
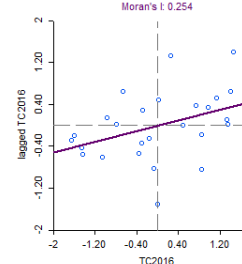
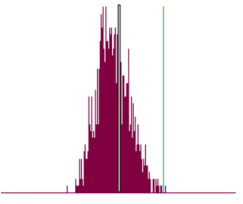
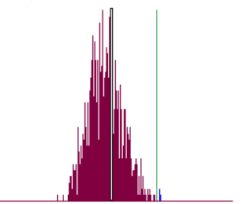
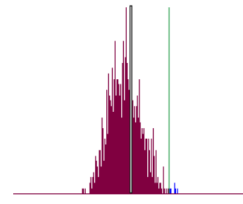
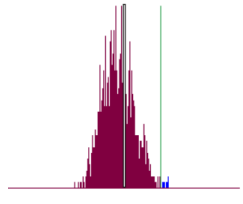
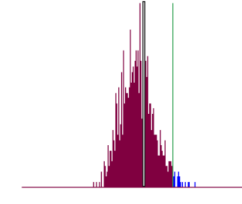

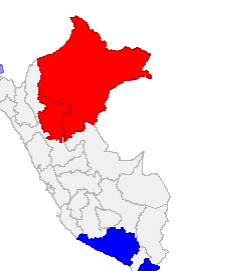



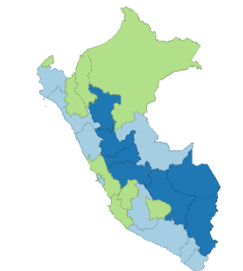

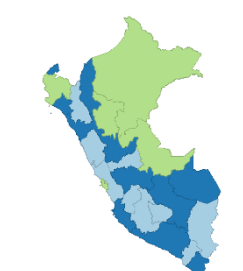
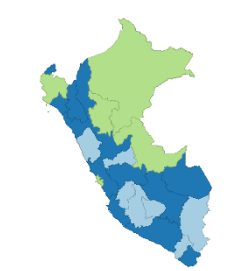
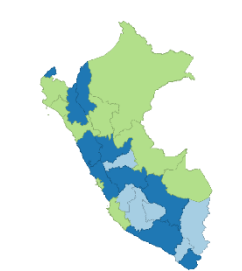
7.5 AEDE del Acceso a Internet.

	2012	2013	2014	2015	2016
<i>I</i> <i>M</i> <i>O</i> <i>R</i> <i>A</i> <i>N</i> <i>G</i> <i>L</i> <i>O</i> <i>B</i>	 <p><i>I</i>: 0.498</p>	 <p><i>I</i>: 0.413</p>	 <p><i>I</i>: 0.440</p>	 <p><i>I</i>: 0.374</p>	 <p><i>I</i>: 0.382</p>
<i>A</i> <i>L</i> <i>E</i> <i>A</i> <i>T</i> .	<p>permutations: 999 pseudo p-value: 0,002000</p> 	<p>permutations: 999 pseudo p-value: 0,002000</p> 	<p>permutations: 999 pseudo p-value: 0,001000</p> 	<p>permutations: 999 pseudo p-value: 0,004000</p> 	<p>permutations: 999 pseudo p-value: 0,003000</p> 
<i>M</i> <i>A</i> <i>P</i> <i>A</i> <i>L</i> <i>I</i> <i>S</i> <i>A</i>					
<i>C</i> <i>L</i> <i>Ú</i> <i>S</i> <i>T</i> <i>E</i> <i>R</i>					

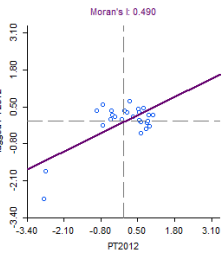
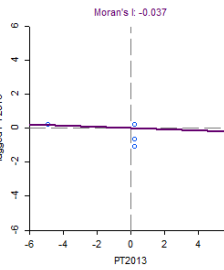
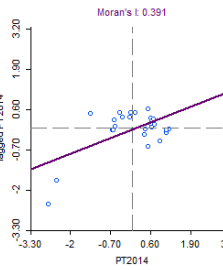
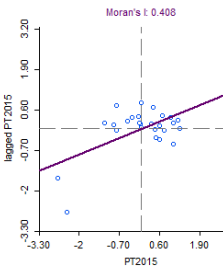
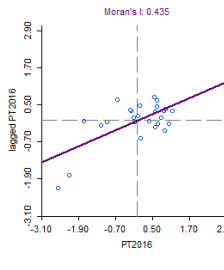
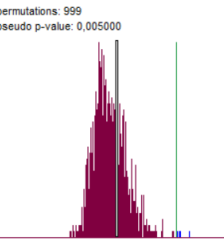
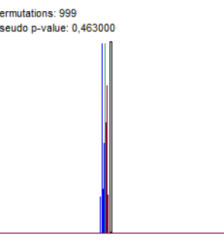
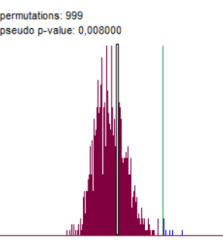
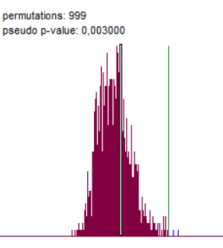
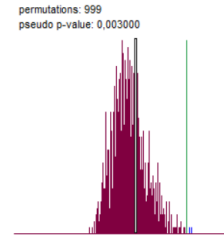

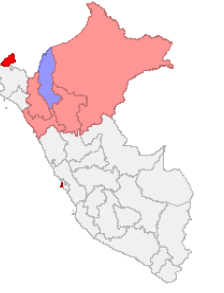



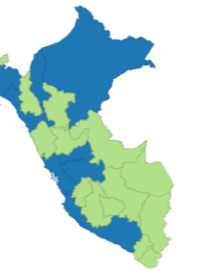
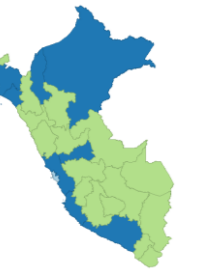


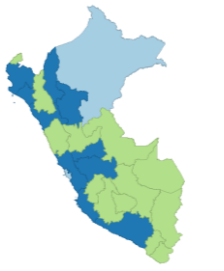
7.6 AEDE de Educación Inicial.

	2012	2013	2014	2015	2016
<i>I</i> <i>M</i> <i>O</i> <i>R</i> <i>A</i> <i>N</i> <i>G</i> <i>L</i> <i>O</i> <i>B</i>	 <p>Moran's I: 0.141</p> <p><i>I</i> : 0.141</p>	 <p>Moran's I: 0.135</p> <p><i>I</i> : 0.135</p>	 <p>Moran's I: 0.153</p> <p><i>I</i> : 0.153</p>	 <p>Moran's I: 0.114</p> <p><i>I</i> : 0.114</p>	 <p>Moran's I: 0.151</p> <p><i>I</i> : 0.151</p>
<i>A</i> <i>L</i> <i>E</i> <i>A</i> <i>T</i> .	<p>permutations: 999 pseudo p-value: 0,093000</p> 	<p>permutations: 999 pseudo p-value: 0,100000</p> 	<p>permutations: 999 pseudo p-value: 0,088000</p> 	<p>permutations: 999 pseudo p-value: 0,136000</p> 	<p>permutations: 999 pseudo p-value: 0,088000</p> 
<i>M</i> <i>A</i> <i>P</i> <i>A</i> <i>L</i> <i>I</i> <i>S</i> <i>A</i>					
<i>C</i> <i>L</i> <i>Ú</i> <i>S</i> <i>T</i> <i>E</i> <i>R</i>					

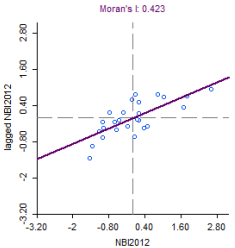
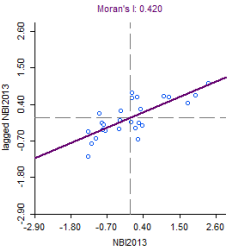
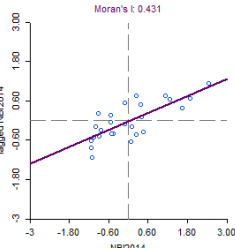
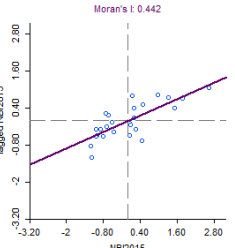
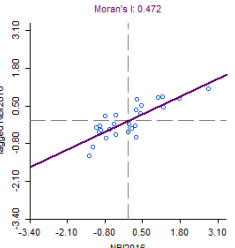
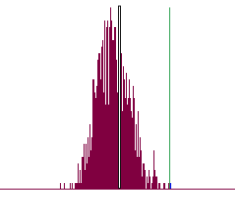
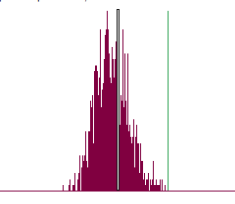
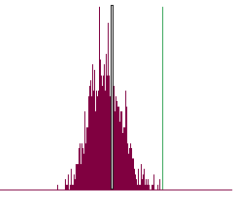
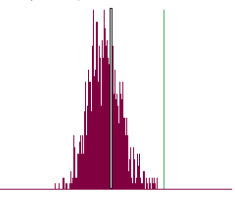
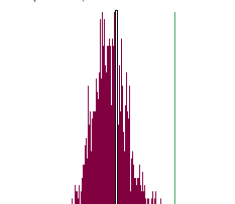










7.7 AEDE de Tamaño de clase.

	2012	2013	2014	2015	2016
I M O R A N	 <p>Moran's I: 0.387</p>	 <p>Moran's I: 0.383</p>	 <p>Moran's I: 0.344</p>	 <p>Moran's I: 0.324</p>	 <p>Moran's I: 0.254</p>
G L O B					
A L E A T.	<p>permutations: 999 pseudo p-value: 0.002000</p> 	<p>permutations: 999 pseudo p-value: 0.004000</p> 	<p>permutations: 999 pseudo p-value: 0.007000</p> 	<p>permutations: 999 pseudo p-value: 0.010000</p> 	<p>permutations: 999 pseudo p-value: 0.020000</p> 
M A P A					
L I S A					
C L Ú S T E R					

7.8 AEDE de Profesores Titulados.

	2012	2013	2014	2015	2016
<i>I M O R A N</i>					
<i>G L O B</i>	<i>I: 0.490</i>	<i>I: -0.037</i>	<i>I: 0.391</i>	<i>I: 0.408</i>	<i>I: 0.435</i>
<i>A L E A T.</i>					
<i>M A P A</i>					
<i>L I S A</i>					
<i>C L Ú S T E R</i>					

7.9 AEDE de Necesidades Básicas Insatisfechas

	2012	2013	2014	2015	2016
<i>I</i> <i>M</i> <i>O</i> <i>R</i> <i>A</i> <i>N</i> <i>G</i> <i>L</i> <i>O</i> <i>B</i>	 <p>Moran's I: 0.423</p>	 <p>Moran's I: 0.420</p>	 <p>Moran's I: 0.431</p>	 <p>Moran's I: 0.442</p>	 <p>Moran's I: 0.472</p>
<i>A</i> <i>L</i> <i>E</i> <i>A</i> <i>T</i>	<p>permutations: 999 pseudo p-value: 0.002000</p> 	<p>permutations: 999 pseudo p-value: 0.001000</p> 	<p>permutations: 999 pseudo p-value: 0.001000</p> 	<p>permutations: 999 pseudo p-value: 0.001000</p> 	<p>permutations: 999 pseudo p-value: 0.001000</p> 
<i>M</i> <i>A</i> <i>P</i> <i>L</i> <i>I</i> <i>S</i> <i>A</i>					
<i>C</i> <i>L</i> <i>Ú</i> <i>S</i> <i>T</i> <i>E</i> <i>R</i>					

7.10 Elección del Modelo General Óptimo.

En primer lugar, se debe identificar si el modelo general presenta problemas de endogeneidad por lo cual se aplica el test de Durbin-Wu-Hausman que compara las estimaciones de la variable instrumental con las estimaciones de mínimos cuadrados ordinarios. Basándose en los resultados de la Tabla N.º 19, se debe aceptar la hipótesis nula que afirma que no hay problema de endogeneidad en el modelo general, por lo tanto, se debe elegir el modelo estimado por MCO.

Tabla N.º 19. Test de Endogeneidad para el Modelo General

Durbin (score) chi2(1)	=	.728617	(p = 0.3933)
Wu-Hausman F(1,122)	=	.687633	(p = 0.4086)

Fuente: STATA

7.10.1 Estrategia: De lo particular hacia lo general.

Anselin (2005) afirma que:

Los test específicos para la autocorrelación espacial sobre los residuos de la estimación MCO, son el I de Moran, y los basados en los multiplicadores de LaGrange, concretamente el LM-LAG, el LM-ERR, el LM-LAG robusto, el LM-ERR robusto. Estos test nos van a dar información sobre la existencia o no de autocorrelación espacial en el modelo y sobre la forma de incorporación de la misma, Si los test no contestan la existencia de ningún tipo de autocorrelación espacial, plantearemos el modelo estimado por MCO (Anselin, 2005).

En caso contrario nos lleva al siguiente paso que es especificar un nuevo modelo que si la considere, estimado por el método de máxima verosimilitud. Para ello se estudian los test más importantes, el LM-LAG y el LM-ERR (Anselin, 2005). Si solo uno de ellos es significativo, escogemos el modelo del error espacial o el del retardo espacial según proceda. Si ambos son significativos, es cuando se debe estudiar los test robustos. Si uno de ellos es significativo escogemos ese criterio para plantear el modelo del error o del retardo correspondiente (Anselin, 2005).

Figura N°4. Proceso de decisión de modelo de regresión espacial



Fuente: Anselin, L. (2005)

7.10.2 Estrategia: De lo general hacia lo particular.

7.10.2.1 Modelo SDM.

En ambos modelos, las variables que tienen una $P > t$ no son estadísticamente significativos y deben excluirse del modelo. En el modelo SDM con efectos fijos, el Rho es igual a 0.69, lo que significa que el rendimiento académico de una región influye en promedio en un 69% a sus regiones vecinas; mientras en el modelo SDM con efectos aleatorios el Rho es igual a 0.60, lo que significa que el rendimiento académico de una región está explicado también por el rendimiento académico de sus vecinos en un 60% (Tabla N.º 20).

Tabla N.º 20. Resultado del modelo SDM fijo y aleatorio

Variable	sdmfe	sdmre
Main		
SB	.00071365	.00092132
ME	-.00019548	.00066197
IN	.00003143	.00034774
ET	.00060836	-.00022672
TC	-.01422226***	-.00640817**
PT	.00037624	-.00073197
NBI	.00037757	-.00088027
_cons		2.6648961***
Wx		
SB	-.00101807	-.00083722
ME	.000682	-.00048976
IN	.00023387	.00040168
ET	.00013141	.00106002*
TC	.00574218	-.00271569
PT	-.00066087	-.00011388
NBI	-.00052537	.00160094
Spatial		
rho	.68967474***	.5988675***
Variance		
sigma2_e	.00015608***	.00025032***
lgt_theta		-1.0900251**
Statistics		
N	130	130

Legend: * p<0.05; ** p<0.01; *** p<0.001

Fuente: STATA

Por otro lado, basándose en el resultado de la prueba de Hausman se debe rechazar la hipótesis nula, por lo cual, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas (Tabla N° 21).

Tabla N°21. Resultados del Test de Hausman en Comprensión Lectora (SDM)

```
Test: Ho: difference in coefficients not systematic

      chi2(16) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =      13.62
      Prob>chi2 =      0.0000
      (V_b-V_B is not positive definite)
```

Fuente: STATA

7.10.2.2 Modelo SEM.

En ambos modelos, las variables que tienen una $P > t$ no son estadísticamente significativos y deben excluirse del modelo. En el modelo SEM con efectos fijos, el λ es igual a 0.75, lo que significa que el rendimiento se explica en un 75% por variables que no están en el modelo de los vecinos mas si en el término error como variables que competen específicamente a la región y que no cambian en el tiempo. Mientras que en el modelo SEM con efectos aleatorios el λ es igual a 0.81 (Tabla N.° 22).

Tabla N.° 22. Resultado del modelo SEM fijo y aleatorio

Variable	semfe	semre
Main		
SB	.0010754	.00088716
ME	-.00055674	.00002944
IN	.00054644	.00070409
EI	.00141282***	.00069226
TC	-.01488939***	-.01028885***
PT	.00011065	-.00049385
NBI	.00036418	-.00049492
_cons		6.4535712***
Spatial		
lambda	.7513989***	.813816***
Variance		
sigma2_e	.00017074***	.00021719***
ln_phi		1.9269196***
Statistics		
N	130	130

legend: * p<0.05; ** p<0.01; *** p<0.001

Fuente: STATA

Por otro lado, basándose en el resultado de la prueba de Hausman se debe rechazar la hipótesis nula, por lo cual, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas (Tabla N.º 23).

Tabla N.º 23. Resultados del Test de Hausman en Comprensión Lectora (SEM)

```
Test: Ho: difference in coefficients not systematic

      chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =      23.40
Prob>chi2 =      0.0015
(V_b-V_B is not positive definite)
```

Fuente: STATA

7.10.2.3 Modelo SDM vs SEM.

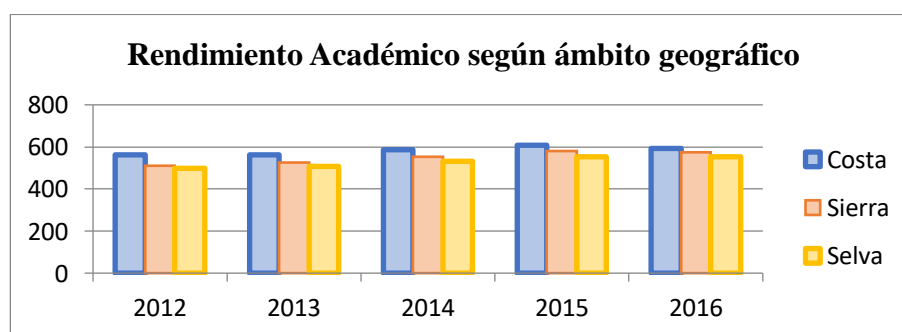
Finalmente se procede a la selección del modelo, mediante el criterio de información bayesiano BIC de los dos modelos de efectos fijos del SDM o del SEM. Por lo general, se prefiere el modelo con el BIC más bajo y en este caso es el del modelo SEM, por lo cual se elige este. Asimismo, se realiza la elección del modelo de regresión espacial mediante el proceso de decisión para elegir el modelo óptimo, del cual se elige y se confirma el usar el modelo SEM para proceder con el análisis de resultados del modelo general.

7.11 Tendencias de las variables según ámbito geográfico.

7.11.1 Tendencia del Rendimiento Académico.

En la Figura N.º 5, se observa la tendencia de los puntajes promedio en Comprensión Lectora durante los años 2012 a 2016, donde la Costa tiene mejor rendimiento académico en Comprensión Lectora que la Sierra y Selva.

Figura N.º 5: Puntaje promedio en Comprensión Lectora, 2012 - 2016

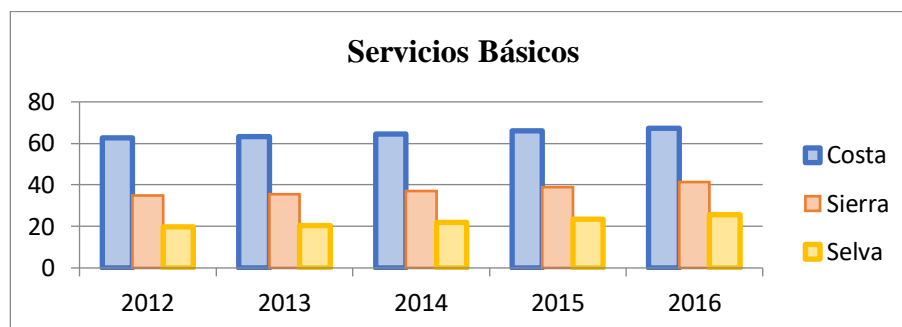


Fuente: SICRECE (2018) - Elaboración Propia

7.11.2 Tendencia de los Servicios Básicos.

En la Figura N.º6, se observa la tendencia de los porcentajes de las escuelas que cuentan con los tres servicios básicos durante los años 2012 a 2016, donde la mayoría de las escuelas que se ubican en la región Costa cuentan con esto servicios, mientras que en la Selva es menor.

Figura N° 6: Porcentaje de escuelas con tres servicios básicos, 2012 - 2016

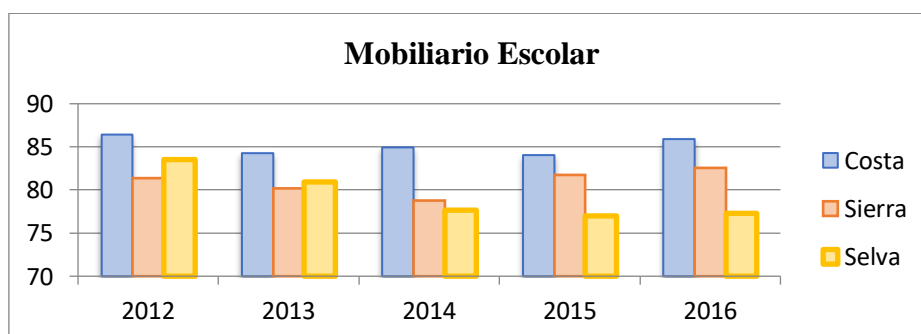


Fuente: ESCALE (2018) - Elaboración Propia

7.11.3 Tendencia del Mobiliario Escolar.

En la Figura N.º7, se observa la tendencia de los porcentajes de escuelas que tienen suficientes carpetas durante los años 2012 a 2016, donde la mayoría de las escuelas que se ubican en la región Costa cuentan con suficiente mobiliario escolar, mientras que en la Selva es menor.

Figura N° 7: Porcentaje de escuelas con suficientes carpetas, 2012-2016

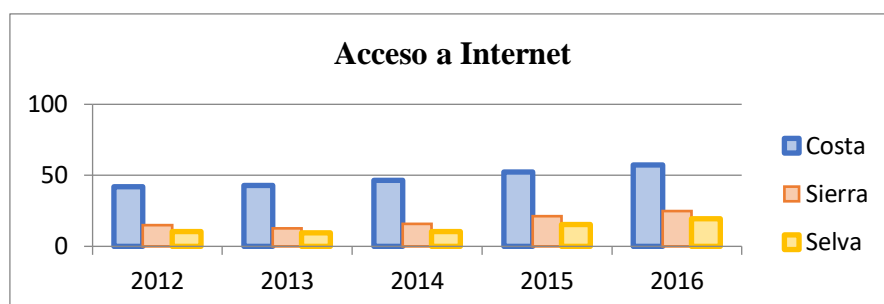


Fuente: ESCALE (2018) - Elaboración Propia

7.11.4 Tendencia del Acceso a Internet.

En la Figura N°8, se observa la tendencia de los porcentajes de escuelas que cuentan con acceso a Internet en primaria durante los años 2012 a 2016, donde la mayoría de las escuelas que se ubican en la Costa tienen acceso a Internet. Asimismo, si bien en la Selva este porcentaje es menor, el acceso a Internet en la región ha ido aumentando lo cual favorece al rendimiento académico de los estudiantes.

N° 8: Porcentaje de escuelas que cuentan con acceso a Internet

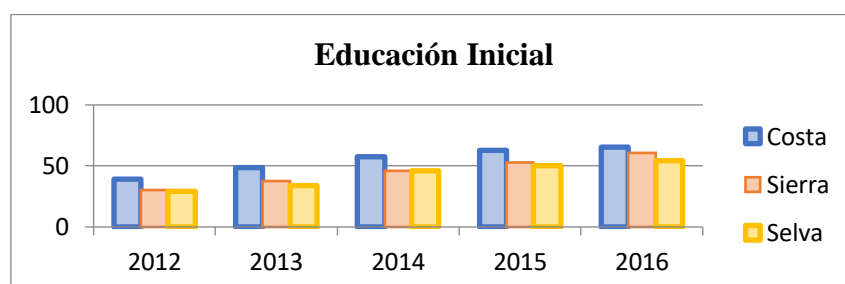


Fuente: ESCALE (2018) - Elaboración Propia

7.11.5 Tendencia de la Educación Inicial.

En la Figura N.º 9, se observa la tendencia de los porcentajes de ingresantes a primaria con tres o más años de educación inicial durante los años 2012 a 2016, donde se ve que se ha ido incrementando en las tres regiones; sin embargo, la región Selva es la que menor porcentaje tienen en comparación que la Costa y Sierra.

Figura N.º 9: Porcentaje de ingresantes a primaria con 3 años de educación inicial, 2012 - 2016

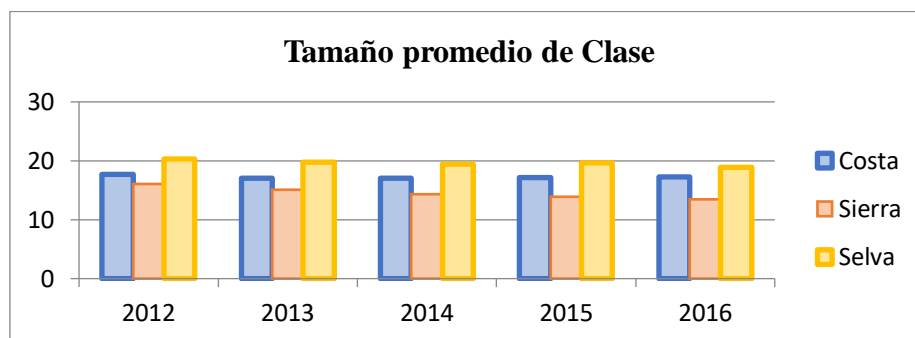


Fuente: ESCALE (2018) - Elaboración Propia

7.11.6 Tendencia del Tamaño de Clase.

En la Figura N.º 10, se observa la tendencia del promedio de tamaño de clase durante los años 2012 a 2016 en las regiones del Perú, donde se ve que en la región Selva, este indicador es mayor.

Figura N.º 10: Tamaño promedio de clase, 2012-2016

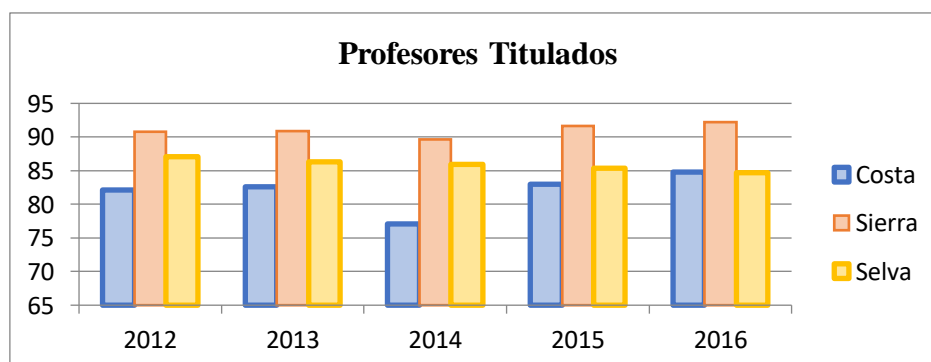


Fuente: ESCALE (2018) - Elaboración Propia

7.11.7 Tendencia de Profesores Titulados.

En la Figura N.º 11, se observa la tendencia del porcentaje de profesores titulados en primaria durante los años 2012 a 2016, donde se ve que en la Sierra hay mayor porcentaje de profesores titulados en comparación con las otras dos regiones.

Figura N.º 11: Porcentaje de profesores titulados en primaria, 2012 - 2016

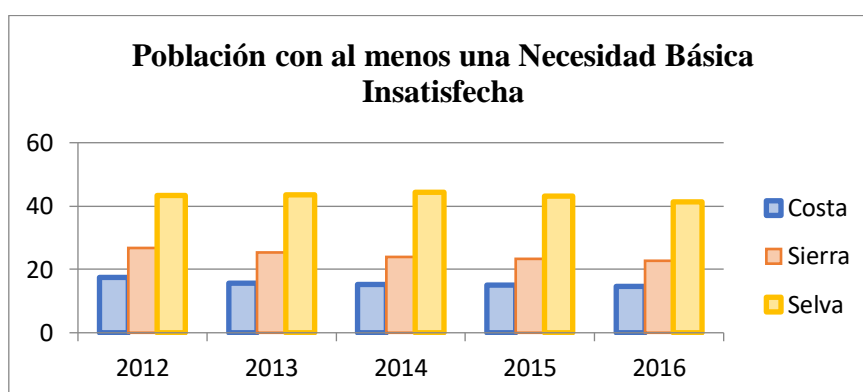


Fuente: ESCALE (2018) - Elaboración Propia

7.11.8 Tendencias de las Necesidades Básicas Insatisfechas.

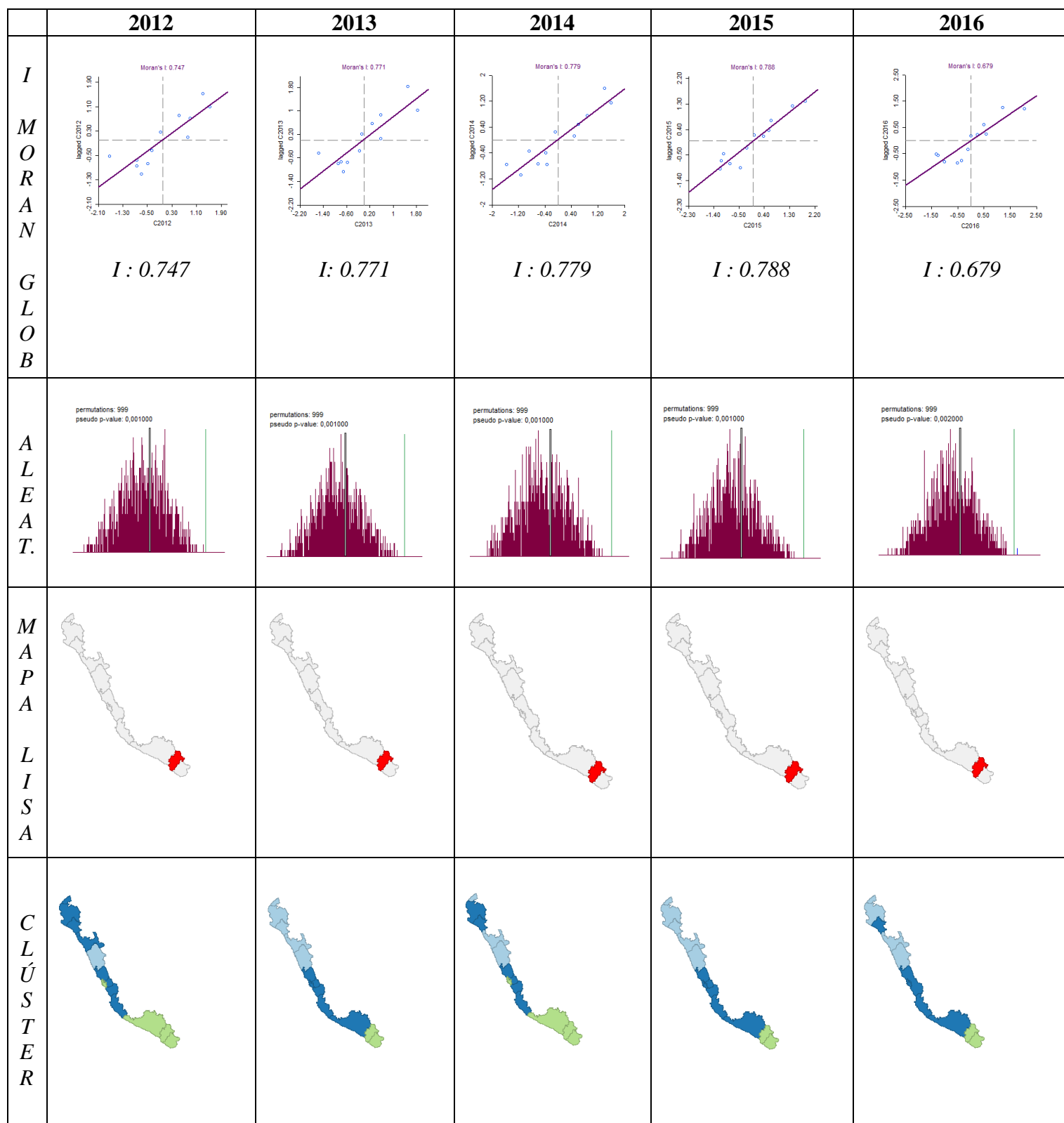
En la Figura N.º 12, se observa la tendencia del porcentaje de población con al menos una NBI durante los años 2012 a 2016, donde se ve que en la Costa existe mayor porcentaje de población que no tiene ninguna NBI; mientras que en la Selva este indicador es mayor.

Figura N° 12: Porcentaje de población con al menos una NBI, 2012 - 2016

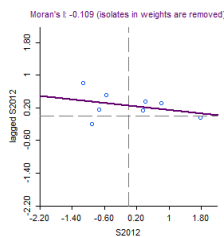
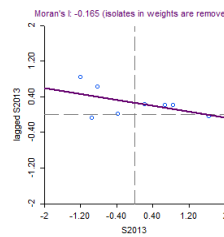

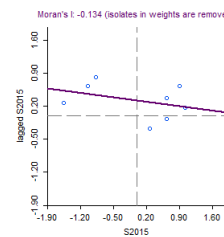
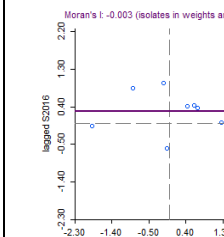
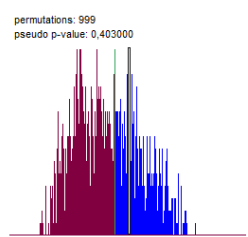
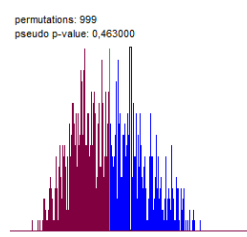
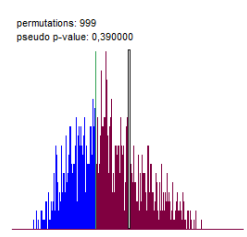
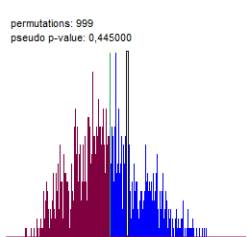
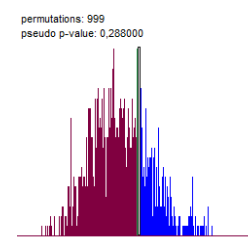





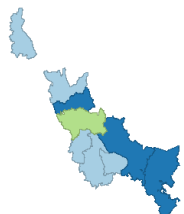
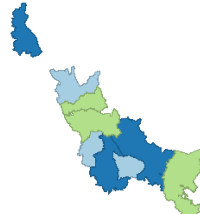
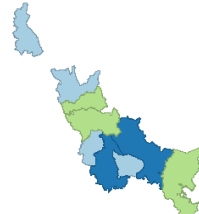
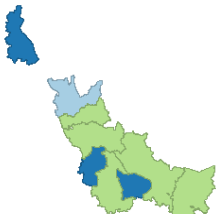
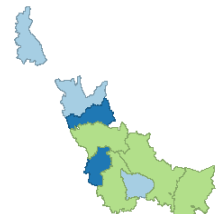


Fuente: INEI (2018) - Elaboración Propia

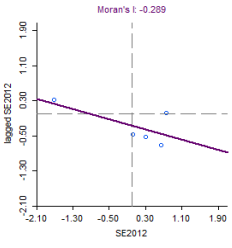
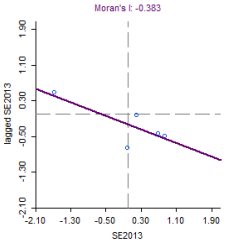
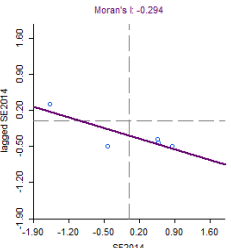
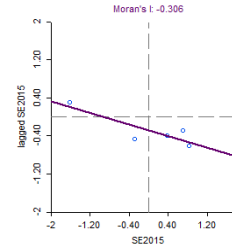
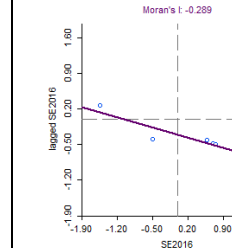
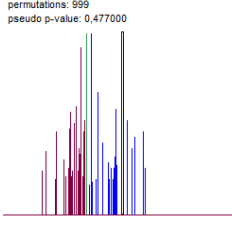
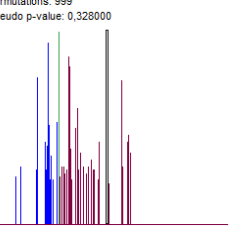
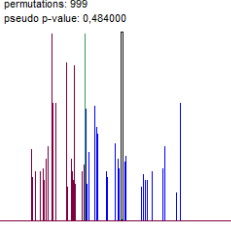
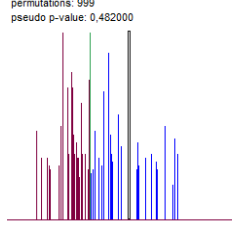
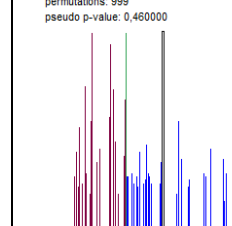





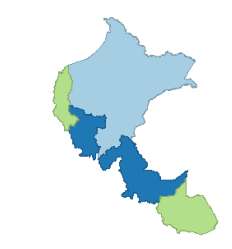
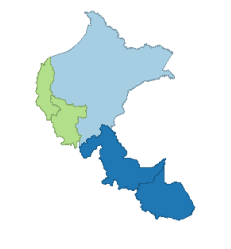


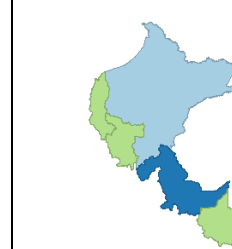
7.12 AEDE del Rendimiento Académico en la Costa.



7.13 AEDE del Rendimiento Académico en la Sierra.

	2012	2013	2014	2015	2016
I M O R A N G L O B	 <p>Moran's I: -0.109 (isolates in weights are removed)</p> <p><i>I: -0.109</i></p>	 <p>Moran's I: -0.165 (isolates in weights are removed)</p> <p><i>I: -0.165</i></p>	 <p>Moran's I: -0.276 (isolates in weights are removed)</p> <p><i>I: -0.276</i></p>	 <p>Moran's I: -0.134 (isolates in weights are removed)</p> <p><i>I: -0.134</i></p>	 <p>Moran's I: -0.003 (isolates in weights are removed)</p> <p><i>I: -0.03</i></p>
A L E A T.	 <p>permutations: 999 pseudo p-value: 0.403000</p>	 <p>permutations: 999 pseudo p-value: 0.463000</p>	 <p>permutations: 999 pseudo p-value: 0.390000</p>	 <p>permutations: 999 pseudo p-value: 0.445000</p>	 <p>permutations: 999 pseudo p-value: 0.288000</p>
M A P A L I S A					
C L Ú S T E R					

7.14 AEDE del Rendimiento Académico en la Selva.

	2012	2013	2014	2015	2016
I M O R A N G L O B	 <p><i>I: -0.289</i></p>	 <p><i>I: -0.383</i></p>	 <p><i>I: -0.294</i></p>	 <p><i>I: -0.306</i></p>	 <p><i>I: -0.289</i></p>
A L E A T.					
M A P A L I S A					
C L Ú S T E R					

7.15 Modelo final para la Costa.

En primer lugar, se debe identificar si el modelo presenta problemas de endogeneidad por lo cual se aplica el test de Durbin-Wu-Hausman que compara las estimaciones de la variable instrumental con las estimaciones de mínimos cuadrados ordinarios. Basándose en los resultados de la Tabla N.º 24, se debe aceptar la hipótesis nula que afirma que no hay problema de endogeneidad en el modelo de la región Costa, por lo tanto, se debe elegir el modelo estimado por MCO.

Tabla N.º 24. Test de Endogeneidad para el Modelo de la región Costa

```
Durbin (score) chi2(1)          = 8.85581 (p = 0.1812)
Wu-Hausman F(1,52)             = 9.004 (p = 0.3298)
```

Fuente: STATA

7.15.1 Modelo SDM.

En ambos modelos, las variables que tienen una $P > t$ no son estadísticamente significativos y deben excluirse del modelo. En el modelo SDM con efectos fijos, el Rho es igual a 0.44, lo que significa que el rendimiento académico de una región influye en promedio en un 44% a sus regiones vecinas; mientras en el modelo SDM con efectos aleatorios el Rho es igual a 0.50, lo que significa que el rendimiento académico de una región esta explicado también por el rendimiento académico de sus vecinos en un 50 % (Tabla N.º 25)

Tabla N.º 25. Resultado del modelo SDM fijo y aleatorio

Variable	sdmfe	sdmre
Main		
SB	.00036507	.00045483
ME	-.00011495	.00081098
IN	-.00033123	.00105989*
EI	.00028316	-.00021495
TC	-.00605007	.00024912
PT	-.00122333	.00075467
NBI	.00035585	-.00232769*
_cons		3.2375379***
Wx		
SB	.00234131	-.00163258
ME	-.00294968*	.00003295
IN	-.00102533	-.00010783
EI	.00178546***	.00065028
TC	.03410003***	.00416073
PT	.00114166	-.00150886
NBI	.00081122	-.00427824**
Spatial		
rho	.44447815***	.49547656***
Variance		
sigma2_e	.00012742***	.00023746***
lgt_theta		-.26249024

Fuente: STATA

Por otro lado, basándose en el resultado de la prueba de Hausman se debe rechazar la hipótesis nula, por lo cual, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas (Tabla N.º 26).

Tabla N.º 26. Resultados del Test de Hausman en Comprensión Lectora (SDM)

```
Test: Ho: difference in coefficients not systematic

      chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =      82.14
Prob>chi2 =      0.0000
(V_b-V_B is not positive definite)
```

Fuente: STATA

7.15.2 Modelo SEM.

En ambos modelos, las variables que tienen una $P > t$ no son estadísticamente significativos y deben excluirse del modelo. En el modelo SEM con efectos fijos, el λ es igual a 0.53, lo que significa que el rendimiento se explica en un 53% por variables que no están en el modelo de los vecinos mas si en el término error como variables que competen específicamente a la región y que no cambian en el tiempo. Mientras que en el modelo SEM con efectos aleatorios el λ es igual a 0.68. (Tabla N.º 27).

Tabla N.º 27. Resultado del modelo SEM fijo y aleatorio

Variable	semfe	semre
Main		
SB	.00127138	-.00007134
ME	.00004215	.00059964
IN	.00131406*	.00142292**
EI	.00083465	.00004132
TC	-.01038552	-.00894003*
PT	.00014756	-.00052802
NBI	-.00057524	-.00153481
_cons		6.4702491***
Spatial		
lambda	.52694235***	.67563331***
Variance		
sigma2_e	.00019461***	.0002358***
ln_phi		1.1222422
Statistics		
N	60	60

legend: * p<0.05; ** p<0.01; *** p<0.001

Fuente: STATA

Por otro lado, basándose en el resultado de la prueba de Hausman se debe rechazar la hipótesis nula, por lo cual, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas (Tabla N.º 28).

Tabla N.º 28. Resultados del Test de Hausman en Comprensión Lectora (SEM)

```
Test: Ho: difference in coefficients not systematic

      chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =      15.47
Prob>chi2 =      0.0304
(V_b-V_B is not positive definite)
```

Fuente: STATA

7.15.3 Modelo SDM vs SEM.

Finalmente se procede a la selección del modelo, mediante el criterio de información bayesiano BIC de los dos modelos de efectos fijos del SDM o del SEM. Por lo general, se prefiere el modelo con el BIC más bajo y en este caso es el del modelo SEM, por lo cual se elige este. Asimismo, se realiza la elección del modelo de regresión espacial mediante el proceso de decisión para elegir el modelo óptimo, del cual se elige y se confirma el usar el modelo SEM en este trabajo de Investigación.

7.16 Elección del modelo óptimo para la Sierra.

En primer lugar, se debe identificar si el modelo presenta problemas de endogeneidad por lo cual se aplica el test de Durbin-Wu-Hausman que compara las estimaciones de la variable instrumental con las estimaciones de mínimos cuadrados ordinarios. Basándose en los resultados de la Tabla N.º 29, se debe aceptar la hipótesis nula que afirma que no hay problema de endogeneidad en el modelo de la región Sierra, por lo tanto, se debe elegir el modelo estimado por MCO.

Tabla N.º 29. Test de Endogeneidad para el Modelo de la región Sierra

```
Durbin (score) chi2(1)          =  .816087  (p = 0.3663)
Wu-Hausman F(1,37)             =  .683399  (p = 0.4137)
```

Fuente: STATA

7.16.1 Modelo aleatorio con efectos fijos y aleatorios.

En ambos modelos, las variables que tienen una $P > t$ no son estadísticamente significativos y deben excluirse del modelo. En el modelo de efectos fijos el Rho es igual a 0.92, lo que significa que el 92% de los efectos fijos individuales lo explica el modelo (Tabla N.º 30).

Tabla N.º 30. Resultado del modelo fijo y aleatorio

Variable	fe	re
SB	-.00350169	-.00086959
ME	-.00079414	.00023158
IN	.00270228*	.0060652***
EI	.00235178*	.00198202***
TC	-.0231684**	-.00774925*
PT	-.00101465	-.01069818***
NBI	-.00121854	.00038761
_cons	6.8149334***	7.2010122***
N	45	45

legend: * p<0.05; ** p<0.01; *** p<0.001

Fuente: STATA

Por otro lado, basándose en el resultado de la prueba de Hausman se debe rechazar la hipótesis nula, por lo cual, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas (Tabla N.º 31).

Tabla N.º 31. Resultados del Test de Hausman en Comprensión Lectora

```
Test: Ho: difference in coefficients not systematic

      chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =      27.72
Prob>chi2 =      0.0002
(V_b-V_B is not positive definite)
```

Fuente: STATA

7.17 Elección del Modelo Óptimo para la Selva.

En primer lugar, se debe identificar si el modelo presenta problemas de endogeneidad por lo cual se aplica el test de Durbin-Wu-Hausman que compara las estimaciones de la variable instrumental con las estimaciones de mínimos cuadrados ordinarios. Basándose en los resultados de la Tabla N.º 32, se debe aceptar la hipótesis nula que afirma que no hay problema de endogeneidad en el modelo de la región Selva, por lo tanto, se debe elegir el modelo estimado por MCO.

Tabla N.º 32. Test de Endogeneidad para el Modelo de la región Selva

```
Durbin (score) chi2(1)          =  4.06591  (p = 0.1167)
Wu-Hausman F(1,32)            =  3.62077  (p = 0.3979)
```

Fuente: STATA

7.17.1 Modelo aleatorio con efectos fijos y aleatorios

En ambos modelos, las variables que tienen una $P > t$ no son estadísticamente significativos y deben excluirse del modelo. En el modelo de efectos fijos Rho es igual a 0.98, lo que significa que el 98% de los efectos fijos individuales lo explica el modelo (Tabla N.º 33).

Tabla N.º 33. Resultado del modelo fijo y aleatorio

Variable	fe	re
SB	.00088413	.00186928**
ME	-.00123641	-.00029077
IN	.00143645	.00228746*
EI	.00334071***	.00271681***
TC	.00582771	-.01195564
PT	-.00164983	.00676758***
NBI	.00568999*	3.218e-06
_cons	5.9578854***	5.7492398***
N	40	40

legend: * p<0.05; ** p<0.01; *** p<0.001

Fuente: STATA

Por otro lado, basándose en el resultado de la prueba de Hausman se debe rechazar la hipótesis nula, por lo cual, se debe elegir el modelo de efectos fijos y eliminar las variables que resultan no significativas (Tabla N.º 34).

Tabla N.º 34. Resultados del Test de Hausman en Comprensión Lectora

```
Test: Ho: difference in coefficients not systematic

      chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =      51.39
Prob>chi2 =      0.0000
(V_b-V_B is not positive definite)
```

Fuente: STATA

7.18 Matriz de Consistencia.

PROBLEMA	OBJETIVO	HIPÓTESIS	VARIABLES	DIMENSIÓN	INDICADOR
<p>General:</p> <p>¿Existe autocorrelación espacial del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 en las regiones del Perú?</p>	<p>General:</p> <p>Determinar la existencia de autocorrelación espacial del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 en las regiones del Perú.</p>	<p>General:</p> <p>Existe autocorrelación espacial positiva del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016 en las regiones del Perú.</p>	<p>V. Dependiente:</p>	<p>Resultado Académico en Comprensión Lectora</p>	<p>Puntaje promedio de la Evaluación ECE en Comprensión Lectora</p>

Específicos:	Específicos:	Específicos:			
¿Cuáles son los determinantes del rendimiento académico en Comprensión Lectora según el ámbito geográfico, durante los años 2012 a 2016?	Identificar los determinantes del rendimiento académico en Comprensión Lectora durante los años 2012 a 2016, según el ámbito geográfico.	Los determinantes que del rendimiento académico en Comprensión Lectora durante los años del 2012 a 2016 son diferentes, según el ámbito geográfico.	V. Independiente:	Servicios Básicos	% escuelas con tres servicios básicos
				Mobiliario Escolar	% escuelas con suficientes carpetas
				Acceso a Internet	% escuelas con acceso a internet
				Educación Inicial	% de ingresantes a primaria con 3 años de educación inicial
				Tamaño de Clase	Tamaño promedio de clase
				Profesores Titulados	% de profesores titulados en primaria
				Necesidades Básicas Insatisfechas	Población con al menos una Necesidad Básica Insatisfecha